APR 29 1924

Journal of Educational Research

OFFICIAL ORGAN OF THE
EDUCATIONAL RESEARCH ASSOCIATION

B. R. BUCKINGHAM, EDITOR

ASSOCIATE EDITORS

E. J. ASHBAUGH — W. W. CHARTERS — GEORGE D. STRAYER
GUY M. WHIPPLE — S. A. COURTIS — LEWIS M. TERMAN
HARRY A. GREENE

VOL. IX APRIL, 1924 No. 4

CONTENTS

	Page
TRAINING STUDENTS IN GROUP INTELLIGENCE TESTING — By Rudolf Pintner	271
THE EDUCATIONAL AND ACCOMPLISHMENT QUOTIENTS AS AN AID IN THE CLASSIFICATION OF PUPILS — By Marvin F. Beeson and Richard E. Tope	281
CHILD ACCOUNTING — By Arthur B. Moehlman	293
THE PRIMARY CLASSIFICATION TEST — By L. W. PRESSEY	305
THE ACCURACY OF CERTAIN STANDARD TESTS FOR SCHOOL CLASSIFICATION — By Percival M. Symonds	315
EDITORIALS	331
REVIEWS AND ABSTRACTS	335
NEWS ITEMS AND COMMUNICATIONS	342
EDUCATIONAL RESEARCH ASSOCIATION	351

Entered as second-class matter at the Post Office at Bloomington, Illinois under the act of March 3, 1879.

Journal of Educational Research

Copyright 1924, by
Public School Publishing Co.

Official Organ of the Educational Research Association

Editor-in-Chief, B. R. Buckingham
Director of Bureau of Educational Research, Ohio State University, Columbus, Ohio

Associate Editors

E. J. ASHBAUGH, *Ohio State University, Columbus.*

W. W. CHARTERS, *University of Pittsburgh, Pittsburgh.*

S. A. COURTIS, *City Schools, Detroit, Mich.*

HARRY A. GREENE, *Ex-Officio, Secretary Educational Research Association, State University of Iowa, Iowa City.*

GEORGE D. STRAYER, *Teachers College, Columbia University, New York City.*

LEWIS M. TERMAN, *Leland Stanford Junior University, Stanford University, California.*

GUY M. WHIPPLE, *University of Michigan, Ann Arbor, Michigan.*

Contributing Editors

MURRAY DALMAN, *Director Department of References and Research, Indianapolis, Indiana.*

VIRGIL E. DICKSON, *Director, Bureau of Research and Guidance, Public Schools, Oakland and Berkeley, Cal.*

CHARLES FORDYCE, *Director of Bureau of Educational Measurements, University of Nebraska.*

WILLIAM S. GREY, *Dean, School of Education, University of Chicago.*

M. E. HAGGERTY, *Dean of the College of Education, University of Minnesota.*

V. A. C. HENMON, *Director, School of Education, University of Wisconsin.*

ARTHUR W. KALLOM, *Assistant Director, Department of Educational Investigation and Measurement, Public Schools, Boston, Mass.*

F. J. KELLY, *Dean Administration, University of Minnesota.*

ROBT. H. LANE, *Assistant Superintendent of Schools Los Angeles, Cal.*

WILLIAM A. McCALL, *Assistant Professor of Education, Teachers' College, Columbia University.*

GEORGE W. MELCHER, *Director, Bureau of Research and Efficiency, Public Schools, Kansas City, Mo.*

ARTHUR B. MOEHLMAN, *Director Statistics and Reference, Public Schools, Detroit, Mich.*

EUGENE A. NIFENECKER, *Assistant Director, Bureau of Reference, Research and Statistics, Public Schools, New York City.*

JOSEPH P. O'HERN, *Assistant Superintendent of Schools, Rochester, N. Y.*

W. J. OSBURN, *Director Educational Measurements, State Department of Public Instruction, Madison, Wis.*

J. B SEARS, *Stanford University, Stanford University, Cal.*

J. L. STENQUIST, *Director Bureau of Educational Research, Public Schools, Baltimore, Md.*

W. W. THEISEN, *Assistant Superintendent of Schools Milwaukee, Wis.*

C. W. WASHBURNE, *Superintendent of Schools, Winnetka, Ill.*

CLIFFORD WOODY, *Director Bureau of Tests and Measurements, University of Michigan.*

The JOURNAL OF EDUCATIONAL RESEARCH is published at Bloomington, Illinois, monthly, except July and August by the Public School Publishing Co. The subscription price is $4.00; $4.20 in Canada or Mexico; and $4.50 in other foreign countries.

The editorial office is at Columbus, Ohio. Address all communications on editorial matters to B. R. Buckingham, Director of Bureau of Educational Research, The Ohio State University, Columbus, Ohio.

Subscriptions and all communications about reprints or other business matters should be addressed to the Public School Publishing Company at Bloomington, Illinois.

JOURNAL *of* EDUCATIONAL RESEARCH

Volume IX APRIL, 1924 Number 4

TRAINING STUDENTS IN GROUP INTELLIGENCE TESTING

RUDOLF PINTNER
Teachers College, Columbia University

Because of the increased use of intelligence tests of all kinds in the schools of today, our colleges of education are finding a greater and greater demand on the part of their students for instruction in intelligence testing. Among those seeking such instruction are experienced teachers who wish to use tests in their classes; teachers who desire to qualify as psychological examiners; supervisors, superintendents, principals, and college professors of education who feel they must keep up with the newer developments in psychology; and students, with no teaching experience, who are interested in psychology and who hope to make use of it later in some way or other. The make-up of a class in intelligence testing at Teachers College, conducted by the writer, was perhaps more heterogeneous than is ordinarily the case at other universities. The great majority of the students were graduate students, with the exception of a few who were within a few points of their B.S. degree. Their previous training in psychology was very diverse, and in the actual subject of testing they ranged from those who had given many tests to those who had never given any. They ranged in chronological age from twenty to fifty. There was a fairly wide range in intelligence scores on such tests as the Otis Higher Examination and the Miller Mental Ability Test. There was also a scattering of foreign students from several different countries. The enrollment of the class was about sixty.

One semester of the work in testing was devoted mainly to group tests, and it is with this part of the work that the present

article deals. Two contrasting methods of instruction might conceivably have been followed. The first method would have been to present the facts about group intelligence tests in a more or less logical order, discuss each group test or type of group test, provide appropriate practice in giving and scoring the test, summarize the results given in the literature about this test, and attempt to evaluate it with reference to its practical value in the school. Logical subdivisions for such a treatment would be: tests for kindergarten and first grade; tests for the primary grades; and so on, up to tests for college students. Another division could be made between verbal and non-verbal tests, and this would lead to a comparison of the two types of tests and their appropriate uses in school. Still another natural division could be made between intelligence tests in the narrow sense, the growing number of tests of mechanical intelligence and business intelligence, and tests of volitional, moral, and character traits.

The second method of conducting the class would have been to train the class in intelligence testing by carrying out real testing in the schools, by actually undertaking the testing required by the schools. This method of conducting the work was in fact followed. Whatever may have been lost in logical presentation and completeness of information about all the different group tests was compensated for by actual practice in giving tests in real school situations and by scoring tests and tabulating results that were going to be used by the schools. This procedure evoked great interest on the part of the students, in all probability more than could have been aroused by a logical and academic treatment of the subject matter.

The general method of procedure was for the instructor to outline to the class the requests from the schools. Usually these came in the form of rather vague demands for some intelligence tests, although sometimes one or more specific tests were mentioned. The class then discussed the intelligence tests which might be suitable, the instructor adding those with which the class was unfamiliar. To take a concrete example, the first

specific task put before the class, shortly after it was organized, was to give group intelligence tests in a high school. This led to a canvass of all available tests. These tests were then administered to the class so that the students might become familiar with their content. The giving of the tests to the class was undertaken by the students themselves, a different student being assigned to act as examiner for each test. After the test had been given, the class criticised freely the procedure of the examiner. Such frank criticism of manner, voice, gesture, and attitude is unquestionably valuable. The test blanks were scored by the class, each member scoring some other student's paper. The papers were then scored a second time by different students. At this point numerous interesting questions were raised, showing the great variation that may exist in the interpretations by different examiners of the scoring directions as given in the test manuals. What probably seemed to the writer of the manual to be "fool-proof" directions turned out in the class to be capable of many interpretations.

As extreme examples of this tendency, we may refer to the general directions for scoring Test I of the Miller Mental Ability Test, which read "mark wrong when the pupil does more than the sentence directs him to do." This direction was interpreted absolutely literally by some scorers, so that for Item 13 where the scoring key gave as the answer "seven" or "7," the scorer marked "seven days" wrong. Again in Item 16, calling for the name of the town or county in which the pupil lives, the scorer marked it wrong if the pupil had added the state,—New York counted right, but New York, N. Y., wrong—because, according to the directions, he had done *more* than the sentence directed him to do. These are extreme examples of the possibilities for divergence in scoring. Many others not so extreme might be quoted.

After the tests had been scored and the methods of scoring had been discussed, one or two students were assigned the task of tabulating the results and of making a report to the class. These reports to the class of the results of the test served as the basis for discussions regarding the various types of statistical

analyses that might be effectively employed. Members of the class unfamiliar with such statistical work found here a potent stimulus for further study.

Having obtained a practical familiarity with the available high-school intelligence tests, the class then discussed those which should be recommended to the school in question, taking into consideration the type of school and the amount of time and money the principal was willing to allow for the testing. In one case this resulted in the recommendation of three group intelligence tests, the results of which were to be combined into some kind of an intelligence index for each child. After the students were assigned for the testing of the various classes, they were met by an assistant and the method of giving the test rehearsed as much as necessary.

The scoring of the first set of test blanks was undertaken by the class as a whole in order to obtain practice in scoring and to achieve uniformity. Later in the semester the scoring was generally undertaken by the group of students who had actually carried out a particular testing program. The latter procedure is preferable after the class has obtained some practice in scoring, because it centers the responsibility more directly. Students are more interested in scoring the papers of the classes they have themselves examined.

At first all papers were checked by another scorer in order to show the students the possibilities of errors that may creep in. After considerable practice in scoring a particular test, differences between the scores of practiced students are small and rescoring becomes unnecessary. The final addition to obtain the total score should always be checked, since there exists here the possibility of a large error.

The first project was reported to the class in great detail. Each student who gave a test and scored the papers for a particular grade presented a distribution of the scores of that grade with the necessary averages and variabilities. Independent treatment of the data was encouraged, so that all possible methods might be demonstrated and discussed. Difficulties encountered in

the classroom while giving the test and difficulties that occurred in scoring the papers were also presented. In subsequent projects this detailed presentation by each student was greatly curtailed.

The separate reports by classes were then assigned to a committee of two or more students—the number varying with the size of the project—so that they might present a final report of the whole school. These final reports afforded a splendid opportunity for showing the originality of the students. The instructor did not specify in detail just what had to be done, he merely gave indications as to possibilities and urged the committee to follow any interesting leads that the data might offer. Many suggestions were thus made for further studies, both by the students presenting the report and by students participating in the discussion of the report before the class.

This final report, after modification in the light of class discussion, was presented to the school. So far as possible this final report was taken by one member of the committee to the principal of the school and explained to her and her teachers in person. This also is excellent training for the student who is to become a school psychologist.

The training in routine testing and the preparation of reports, although valuable in themselves, are not, however, the only advantages accruing from this practical method of conducting such a class. Fully as important is the fact that this procedure naturally encourages the student to raise real questions with reference to any particular group of children tested. An intelligence survey of a school is bound to suggest other problems for investigation. A school made up largely of children of foreign-language-speaking parents led the class into a discussion of the value of verbal and non-verbal tests. This resulted in the giving of one test of each type and in comparing the results. In every school interest is bound to center on individual children who make extremely high or extremely low scores. In one school several such cases were investigated, one case being assigned to each student who volunteered to make a detailed study. A somewhat abbreviated report on a special case prepared by Miss Ruth Andrus, a student in the class in mental tests, is included here.

Case Study

Name: Henry S.
Date of Birth: Dec. 20, 1915
Race: White

Preliminary Statement of the Problem.—On November 21, 1922, the students in Education 451, Teachers College, made a survey of the school which Henry attends. In grades I and II the Kingsbury and Pressey Primary Tests were given. From the results of these two tests it was reported that Henry had an I.Q. of 120 and the request was made that this case be the subject of special study.

Neighborhood.—The neighborhood is known as Hell's Kitchen and as an environment leaves much to be desired. The house in which the home is located is one of the best on the block; the hallways are fairly light and quite clean. Moreover, the building seems to be in comparatively good repair. The S. family live in the flat third-floor back.

Family History.—Henry's father is a large, good-natured man, easy-going and contented with little; he does not like indoor work so drives a truck on the docks. He finished school at thirteen, having completed the seventh grade. He was born in the neighborhood in which he now lives and went to the same school which Henry now attends. His father came from Germany, when a child, and the family settled in this same neighborhood. Henry's paternal grandmother was born in New York a little farther downtown than their present home. Both the paternal grandparents are living and in good health.

The mother's family for two generations have lived in this same neighborhood. She is a wiry, thin woman who says she has always been "bronical" (bronchial). She did not go to school until she was nine because she lived with her grandmother, who "spoiled" her, but she had nearly finished the seventh grade when she had to go to work at fourteen. Her father was an iron-moulder and died, when she was very young, of typhoid fever.

From both the paternal and maternal sides, there seems to be a generally good inheritance but not sufficient ability to rise above the environment in which they found themselves.

Early History of Henry.—Henry was a bottle-fed baby, and had four teeth at seven months. He began to walk at sixteen months, but the date of his beginning to talk is uncertain. He was a good baby, not "colicky." He has had measles, chicken-pox, and a slight attack of diphtheria. He has never had convulsions and has no symptoms of heart trouble. He has some bronchial disorder and has had an abscess in one ear. He is troubled with enuresis even yet, when sick. Henry learned his letters and some words before he went to school because "he's always askin' questions and of course I have to answer him," says his mother.

Mrs. S. has had three children of whom Henry is the youngest. The oldest is a boy of sixteen who works now, because they could not persuade him to go to high school, although he finished the eighth grade

at fourteen, having skipped a grade. There was a little girl who died of diphtheria at two and one half years. Perhaps this child's death explains the extreme watchfulness and worrying over Henry. Mrs. S. is always running to the doctor with him. She goes to school after him in the afternoon, takes him for a walk, watches from the doorway when he plays with the children, and on Sunday gives up her day to him, taking him out walking to his grandfather's, etc. Both boys go to a Congregational Sunday School in the neighborhood regularly. The mother has good ideals, watches over her children, and does her best to make her home what it should be. The teachers and the principal said that Henry's physical welfare was looked after better than that of any other child in the school. His teachers speak of him as an unusual child and the joy of their lives. "He would stand out anywhere but specially here."

Tests.—By the Stanford-Binet scale Henry's mental age is nine years with an I.Q. of 124. The basal year is five. In Year VI he missed "mutilated pictures." He passed all of VII; all but "ball and field" in VIII; all of IX; and two, "absurdities" and "comprehension," in X; but none in XII and XIV. In the case of the "reading" and "problems" in X, the failure was due to lack of time as he reads well but slowly, and his method in the problems was careful and accurate but slow. Of the "designs," he produced one with full credit but did not look at the other, so failed it entirely. When I questioned him about this neglect, he said, "I don't know why. I just looked at one." In this case he showed no apparent sense of failure. An echo from his environment appears in Test 4-b in XIV, for when he said, "some one died" and I asked him why the lawyer had come, he said, "to see if the man had ever been in jail." On the whole, he seems to have more ability with words than with symbols, either numbers or pictures, although he says he likes arithmetic.

In the Performance Scale his M.A. was 9-5 and his I.Q. was 129. In the Ship Test and Picture Completion, he showed the same lack of discrimination as in the "mutilated pictures" in the Stanford. His M.A. on those two tests was 6 years and 7 years. On the Herring revision of the Binet, his M.A. was 9-3 and his I.Q., 127.

I also gave Dr. Woodworth's test, a questionnaire to detect neurotic tendencies in children. In this test he appeared under- rather than over-emotional. He showed a lack of imagination and a strong tendency to accept conditions without question which showed itself in a lack of discrimination in the first two tests given. This appeared especially when he was asked if he ever made believe that the things he wished for were true. He said, "No," and when asked why, he said that he had everything that he wanted. For a child with his intelligence quotient and in his environment that was an astonishing admission. While he is alert, this alertness is confined within narrow limits forced upon him by his environment. This is seen in the Picture Completion Test in which he placed the "broken window" correctly, but put a "clock" where

the "log" should be and a "sleeping cat" where the "bird" should be flying from the "cage." There were other glaring inconsistencies. He showed some ability to criticise his results. In this connection, it is interesting to note that his responses to and interpretation of the "absurdities" in X (Binet) were quick and accurate. Pictures and judgment about them seem to lie outside of his experience.

Henry does not seem puffed up at his school success. The fact that he is the honor pupil of his class month after month does not affect him. He goes his own quiet unenthusiastic way, "thinking about his school work, though it isn't hard," playing with the boys, wishing his mother would not come to school for him or watch him from the doorway, because then the children don't like to play with him. He does not know what real joyful play means, however, and there is a lack of enthusiasm about him which may be the same thing as the lack of ambition seen in his father and brother.

Throughout the testing, Henry seemed at his ease and eager to do his best but was rather easily discouraged and very acquiescent in his general attitude. He was cooperative but absolutely lacking in enthusiasm though he said he liked the tests, especially the Performance Scale. His teachers also mentioned this lack of enthusiasm and his acceptance of conditions without question. Because this attitude seemed to be combined with a lack of persistence and of discontent with his environment, he furnished a good example of the fact that the sort of intelligence indicated by the intelligence quotient does not tell the whole story. The intelligence quotient must be combined with emotional strength or force or interest, or what you will, so that a certain divine discontent may be present and spur the person on. Henry's father and brother showed this lack of ambition as have the other members of the connection, and one wonders what will become of Henry with his I.Q. of 124 and his placid acceptance of things as they are.

Conclusions.—To broaden Henry's interests and arouse enthusiasm and imagination, I would suggest that he be placed in a Special Opportunity Class for Bright Children, if such a class is formed, and that the curriculum of this class be enriched rather than merely accelerated. Any healthy play interests which can be developed in connection with such a class would be of inestimable value, and also any activity which will develop Henry's initiative in either play or school work. The trips which are planned for the classes in this school will be a very fine thing for a child like Henry. Creation of enthusiasm and development of interest and initiative are necessary if Henry is to realize the worth of his intelligence. If the principal thinks it advisable, it might be helpful to explain the situation to Mrs. S. and enlist her aid, for she seems capable of appreciating situations presented to her, and Henry is her main interest.

This study of a particular case is one of several undertaken by various students. Such studies gave the class some insight

into what is meant by clinical psychology and helped them to a better understanding of the position of the intelligence test in the attempt to get a picture of the whole personality of any individual child.

As a sample of a different type of study undertaken may be mentioned the report of Mrs. A. H. Talbot, a member of the class, on the comparison of the intelligence quotients of siblings with those of non-siblings. She found the following relationships between the intelligence quotients:

Grades	Siblings		Non-Siblings	
	Number of Pairs	Correlation	Number	Correlation
I, II	34	+.42	34	+.14
III, IV	19	+.24	19	−.03
I to IV	53	+.33	53	+.09

The tests given in grades I and II were different from those given in grades III and IV.

Other studies undertaken by the class involved a survey of a rural district, group tests in a deaf school, the giving of a battery of educational and intelligence tests to all chronologically overage children in several large schools in the city, and so forth. Although no one student was able to take part in every project, the number and diversity of projects were so great as to insure to each student many different types of practice in group testing.

In summing up the work of this class in group intelligence testing, it is the opinion of the writer that this problem or project method of procedure is admirably adapted to the training of mature graduate students. It may lack in logical presentation and equally distributed practice among all students in all tests. From the very nature of the work some particular tests were stressed more than others, because they were suitable for the actual jobs that had to be handled. Furthermore, some students received much practice in giving and scoring certain tests, while others received relatively little with those same tests. Some useful intelligence tests were not used at all by the class. Such unevenness and such omissions are, I believe, more than adequately compensated for by the really intimate knowledge

obtained from several tests functioning in real school situations. The value of group tests in the schools, their advantages and drawbacks, can best be seen by students who are doing the work themselves. Needless to say, the interest on the part of the student in such work is great.

This method of training in group testing offers one other advantage to the instructor, who is required to recommend his students for positions as school psychologists, mental examiners, and the like. He is afforded the opportunity of watching the student in a real school, of noting his attitude before a class and his handling of the classroom teacher. Giving a group test to a group of fellow-students in the laboratory is very different from giving a test to a crowded room of fifty or sixty youngsters with, perhaps, a fussy or hostile teacher in the room. And, finally, the preparation of the report gives the instructor some insight as to the accuracy of the student and his appreciation of the significance of his test results.

THE EDUCATIONAL AND ACCOMPLISHMENT QUOTIENTS AS AN AID IN THE CLASSIFICATION OF PUPILS

MARVIN F. BEESON
Formerly Director of Cooperative Extension Service, Western Colorado
AND RICHARD E. TOPE
Superintendent of Schools, Grand Junction, Colorado[1]

Various plans for classification have been suggested as substitutes for the unsatisfactory method of placing pupils in grades or rooms on the basis of chronological age, number of years in school, the completion of certain books or courses of study, and the general quality of work as judged by the teacher. Under the customary plan the assumption is that all the pupils of a room are approximately equal in ability.

It is now well known that this is not the case. The range of mental ages in nearly every grade varies as much as seven years. McCall found in a survey of two typical schools, reported in his *How to Measure in Education*, the following results: (1) about 44 percent of pupils were wrongly classified; (2) about 34 percent of pupils were misplaced one grade; (3) about 10 percent of pupils were misplaced two or more grades; and (4) only about 8 percent of pupils were pushed ahead of the grade where they belonged, while nearly 36 percent were held back from the grade where they belonged.[2]

Kelly, Ruch, and Terman state that in the four California school systems upon which the norms for the Stanford Achievement Tests were based, approximately 65 percent of the 12-year-old pupils were at that time in the wrong grade. This brings about a retardation of the gifted pupils and an acceleration of

[1] This study was prepared as a part of a seminar course in educational research conducted by the Cooperative Extension Service of Colorado during the session of 1922-23. A large part of the work was done by the following committee of teachers in the Grand Junction Schools: Mabel Hoffmann, Mary Rait, Lenore Watkins, Frances Strock, Mabyl Righdenour, Doris Phenice, and Olive Jorgensen.

[2] McCall, W. A. *How to Measure in Education.* New York, Macmillan Company, 1922. p. 52.

the dull. As Franzen has expressed it, "teaching in such grades simply mystifies the lower quarter and bores the upper quarter."

According to McCall, the two chief objectives of classification are to group the pupils according to equal educational status and according to equal ability to progress. The real difficulty, then, is to find a method of classification that will best provide for both these objectives.[3]

Grouping by mental age and intelligence quotient ignores the first of these objectives, but provides in part for the second. Certain factors, such as health, industry, personal traits, varying tastes and interests, home environment and social status, and the culture and attitude of the parents—influences which affect a child's ability to progress—are not covered by this method of grouping. Practically all of these influences are positively correlated, but the correlations are not very high, and even if they were the rule would not apply to all individual cases. Perhaps we cannot hope to find a method of classification that will provide for all factors.

Another deficiency in mental tests as sole criteria for classification and promotion is that they do not take into consideration the educational status of the child. They probably do this indirectly and to some extent; but children who are classified or promoted by intelligence tests alone are likely to have some gaps in their educational careers. These are not so disastrous as was formerly thought, for it has been shown that the bright child rather quickly and easily fills in these gaps because of his superior intelligence. We cannot believe, however, that this is the best method of progress. Surely some more uniform and continuous method would be preferable, for ability to progress will not always insure that the children will actually make all the progress of which they are capable. Therefore it is best that the intelligence tests should be supplemented by other criteria.

In the schools of Grand Junction, Colorado, it has been the custom for four years to divide the pupils of the sixth, seventh, eighth, ninth, and tenth grades into homogeneous groups. By

[3] *Ibid.* p. 39-40.

this method pupils are put together who are nearly equal in educational status, and who can or will progress at about equal rate. This homogeneous grouping is based upon intelligence quotients. The classification is based upon the National Intelligence Tests for the sixth grade and upon Terman's Group Test of Mental Ability for grades above the sixth. The results of the intelligence tests are not followed rigidly, but health and such personal traits as effort and industry are also considered.

The grades are divided into five sections. The grouping is determined by the intelligence test and the judgment of the teacher. Educational tests, given later in the year, serve as a basis for some regrouping. Improvement in the individual pupil's work also leads to readjustment. In the eighth grade, where the grouping is perhaps most complete, changes are made at the end of every six weeks. When a pupil improves his work sufficiently, he is advanced to a higher division. The range of intelligence quotients in the best division of the present eighth grade is from 94 to 148, with a median of 132. In the poorest division the intelligence quotients range from 73 to 104 with a median of 91. Probably an improvement over the present system would be to regrade the pupils on the basis of mental age, and arrange them in sections according to the intelligence quotients.

The problem is somewhat different in the high school, since the course of study allows the pupils a choice. The same plan of grouping is followed in general, however; and, especially in the required subjects, a fairly homogeneous grouping is obtained. New pupils entering the schools are required to take an intelligence test at once, and on this basis alone they are placed in the group to which they appear to be best suited. The teachers find that when the pupils are grouped according to intelligence, the real teaching problem has become that of adapting the work and the methods to the ability of the group. The problems of failure and promotion have been practically eliminated, and the pupils advanced from year to year have continued to work more nearly up to their capacity. The greatest and most important

problem of all has been that of planning the course of study in such a way that appropriate subject matter and special or miscellaneous assignments were arranged in the proper sequence, so that educational progress was continuous, and at the same time suited to the ability of all the groups.

That the plan has proved successful can hardly be questioned. It has made possible the adaptation of instruction to the needs of the pupils in a way that was entirely impossible under the old system. The pupils grouped according to intelligence are, with a few exceptions and adjustments, also grouped as to temperament, personal traits, and physical development. The same appeal, therefore, brings a similar reaction throughout a section, whereas an entirely different approach may be necessary with another group.

In classifying children in the primary grades the mental ages and intelligence quotients are probably the best bases for classification, since there are few reliable educational tests for these ages. A similar plan should be followed for the grouping of high-school pupils because of the great variety of subjects taught in the high school and the few really standardized tests available for those subjects.

In the intermediate and upper grades probably the best method of classification is by means of the educational age (E.A.) and the educational quotient (E.Q.). At present the accurate determination of the educational age is difficult since few of the standardized tests give age norms. The custom of giving age norms, in addition to the grade standards, for the educational tests is, however, becoming much more common.

The educational quotient is the relation of the educational age to the chronological age, and reveals whether the pupil's educational attainment is above or below the typical achievement for his years. It should normally approximate 100.

Classification by educational age meets the first objective already mentioned, for it brings together pupils of equal educational status. Educational age shows directly which pupils are of equal educational attainment, while mental age does this only indirectly. The educational age is in reality, a resultant

of mental age, health, industry, attendance, attitude toward school work, and various other important influences.

The second criterion in classification, that of bringing together pupils who will progress at an equal rate, is taken into consideration by the use of the educational quotient. The best way to determine what the rate of progress will be is to find out what it has been. The mental age and educational age, considered apart from chronological age, reveal practically nothing about the past rate of progress. This rate of progress is shown for intelligence by the intelligence quotient, but for education it is shown more directly by the educational quotient. If a pupil's education progress has been rapid, his educational quotient will be proportionally high. If it has been slow, the educational quotient will be lower.

Reliance upon the educational tests alone is unfair to the pupils, because high rank and praise are thus given to the pupils who score above the norm and failure and blame to those who score below the norm, without regard to the native capacity or intelligence of the children. Test scores are unfair to the teacher, also, because they show only how her grade compares in measurable educational abilities with other grades in the country without considering the pupil-material with which she works. Those interested in educational tests have long realized that both the teacher and pupil should be measured, not only by what the child has learned, but by what he has learned in relation to what he is capable of learning. In order to get the results of these mental and educational tests in one measure and in terms that all may comprehend, the accomplishment quotient or achievement quotient has been devised.

The accomplishment quotient (A.Q.) is the ratio of the educational quotient to the intelligence quotient, or the educational age to the mental age. The formula given below will clarify this.

$$AQ = \frac{EQ}{IQ} = \frac{\frac{EA}{CA}}{\frac{MA}{CA}} = \frac{EA}{MA}$$

The accomplishment quotient is a summary of what the child has accomplished educationally compared with what he is capable of accomplishing. It is, therefore, the fairest and most valuable measure of the efficiency of both teacher and pupil, which has yet been devised.

For the teacher the accomplishment quotient is the only quotient which takes into account the material with which she is working. A knowledge of the accomplishment quotients of her pupils is a protection against an unjust evaluation of the work of the teacher. The chief criterion for rating a teacher should be the results achieved, particularly if the mental ability of the pupils is considered. This appears to be a much more direct method of estimating a teacher's efficiency than by such qualities as handwriting, voice, appearance, poise, optimism, health, experience, preparation, and professional study, especially since Knight[4] has shown in a recent study that none of these qualities is highly correlated with teaching efficiency.

For the pupils the accomplishment quotient shows who are living up to their possibilities and who fail to make attainment equal to their capacity; who are deserving of praise, and who merit censure. With this criterion in mind it is expected that the child will progress according to his innate mental capacity, which is the only fair standard.

Thus it follows that the incentive yielded by the use of accomplishment quotient is one of its greatest values. Toops and Symonds,[5] to whom we are indebted for several of the ideas in this report, point out that pupils by graphing their improvement in accomplishment quotients and by receiving special educational treatment have brought their accomplishment quotients up to 100.

The accomplishment-quotient procedure is valuable also as an element in school placement. By its use schools may provide individual rates of progress for every pupil, such as are now often attempted in certain business colleges and trade schools.

---

[4] Knight, F. B. "Qualities Related to Success in Elementary School Teaching." *Journal of Educational Research*, 5:207-16. March, 1922.

[5] "What Shall We Expect of the A.Q.?" *Journal of Educational Psychology*, 13:522, December, 1922.

With rapid promotions based on ability we shall approach perfect correlation between intelligence and school ability. Some educators, however, do not approve of rapid promotion, but advocate instead an expansion of the topics to be covered by the superior pupils.

The accomplishment quotient does not determine the cause of poor work, however; it merely singles out the pupil who needs diagnosis and treatment. Pintner and Franzen recommend its use during the school year to discover those pupils, those classes, and those grades or schools that are not working seriously enough. The discovery of the cause and remedy for such conditions is, of course, another problem.

In the words of Toops and Symonds, "The need in education for a measure of accomplishment, achievement, effort, motivation, attitude, dynamic effect of educational environment —call it what you will—is so great that it seems quite possible that the accomplishment-quotient procedure may be accepted uncritically by educators."[6] If, however, the accomplishment-quotient method brings the increased flexibility of promotion shown to be desirable, if it results in newer and increased emphasis upon more adequate sectioning at the beginning of each semester, if it makes more evident that a child newly promoted may be deficient in the tools of education at that point, if it provides the real incentive which it promises for making indolent pupils work, then it is a real and worthy object of admiration.

The use of the accomplishment quotient is one of the most recent developments of educational psychology; consequently not all authorities agree as to its technique and interpretation. According to Franzen,[7] an A.Q. of 100 indicates what a pupil is able to do under the best conditions, and according to Monroe and Buckingham,[8] who use the term "achievement quotient," it means that the pupil has achieved exactly as much as the average

[6] "What Shall We Expect of the A.Q.?" *Journal of Educational Psychology*, 14:27-38, January, 1923.

[7] Franzen, Raymond H. "The Accomplishment Quotient." *Teachers College Record*, 23:434, November, 1920.

[8] *Teacher's Handbook, Illinois Examination I and II.* Bloomington, Illinois, The Public School Publishing Company, 1922. p. 16.

of the pupils of his mental age. There are further difficulties involved in the terminology used, in the norms employed, in the selection of standard tests, and in securing alternative tests of capacity and attainment.

There are other more serious drawbacks to the use of the accomplishment quotient in the present stage of the evolution of standard tests. In the first place, for many of the achievement tests, as has been said, age norms have not been determined; and this necessitates a very approximate method of computing the educational age, educational quotients, and accomplishment quotients. In the second place, the standards on different tests differ to such an extent that the resulting educational quotients, intelligence quotients, and accomplishment quotients are not comparable. Furthermore, there is the embarrassing fact that different achievement tests even of the same subject measure different aspects of the subject, and measure them in a different way, so that the average intercorrelation between different standard tests of the same subject seems to be only from about 0.40 to 0.60, as Henmon has recently shown. He states that "the intercorrelations between the different *informational tests* in history range from 0.24 to 0.82, with an average of 0.58 for sixteen coefficients." The inter-correlations for comprehension of reading range from 0.02 to 0.77 with an average for twenty coefficients of 0.40. In speed of reading the correlations are naturally higher. They range from 0.38 to 0.88 with an average of 0.60 for twenty coefficients.[9]

Similar but less marked inconsistencies arise in the case of mental tests. Strictly speaking, the intelligence quotient refers only to the results obtained from the Stanford Revision of the Binet-Simon Scale, and the intelligence-quotient equivalents for the group tests of mental ability are only approximate. Hence we are dealing with two variables, the variability in the achievement tests and the variability in the intelligence tests.

There is still another danger involved in the computation of the accomplishment quotient. A single application of even

[9] Henmon, V. A. C. "Some Limitations of Educational Tests." *Journal of Educational Research*, 7:185-98. March, 1923.

a very excellent mental or educational test will give only approximate results. The very least we should do, therefore, is to attempt to ascertain the intelligence quotient accurately, preferably by means of the Binet scale, or at least by the repetition of two or more forms of a group test of intelligence. Similar precautions should be taken with regard to the achievement tests. It would probably be best to give two or three different tests for each school subject included in the computation of the educational age and educational quotient, rather than to use several forms of the same test. In this way several different aspects of each school subject may be measured, and errors in the norms may be offset.

If regular age standards for a test are lacking, they may be computed approximately according to Table I, which is a modified form of a table worked out by McCall.

TABLE I. AGE STANDARDS

Grade	September Age
I	6–8
II	7–9
III	8–10
IV	9–11
V	11–0
VI	12–1
VII	13–2
VIII	14–3
IX	15–4
X	16–5
XI	17–6
XII	18–7

The accuracy of the computations can, of course, be increased by interpolating in the scores the differences from the ages indicated above. This method of determining age norms is very unsatisfactory, however, so that if possible tests should be used for which accurate age standards, as well as grade standards, are available.

The most suitable battery of achievement tests at present available for this purpose seems to be the Stanford Achievement Tests. The aim of the authors, as the manual for these tests

indicates, was to construct tests which would cover practically all the curriculum from grades II to VIII, which would yield consistent results (have high reliability), and which would have the greatest possible validity as a basis for the grading and classification of pupils. The tests contain only those subjects deemed of most importance and those that are likely to be covered in all schools.

The series includes a threefold test in reading which is designed to cover the three main factors involved in getting the meaning from the printed page: knowledge of word meaning, ability to comprehend relatively short language units (sentences), and ability to grasp the meaning of larger units (paragraphs). In arithmetic there are tests in the fundamental operations and tests in problem solving. The other tests included in the series are: nature study and science combined, history and literature combined, language usage, and a dictation exercise for spelling.

The time required for the primary examination is about 70 minutes, and for the advanced examination about 135 minutes. This is probably long enough to give a rather accurate indication of the knowledge of the subjects included in the series. The sittings may be held on one day or several days.

The scoring is made objective by a very convenient system of keys. The scores are expressed in points, and a method is given by which the points may be converted into subject ages for each subject and into educational ages for the whole series. It will be observed that many of the criticisms of the accomplishment quotient cited above are met by this carefully devised and valuable series of achievement tests.

In the Grand Junction schools a brief study was made of the intelligence quotients, educational quotients, and accomplishment quotients of one hundred unselected seventh-grade pupils, based upon the Terman Group Test of Mental Ability and the Stanford Achievement Tests. The exact chronological age of each pupil was found, and from these tests the intelligence quotients, educational ages, educational quotients, and accomplishment quotients were computed.

An examination of the results of the 100 tests discloses the fact that those pupils whose intelligence quotients were above 100 almost invariably have lower accomplishment quotients while those whose intelligence quotients fall below 100 have higher accomplishment quotients. In the hundred cases only three pupils whose intelligence quotients were above 100 made higher accomplishment quotients, and only three pupils whose intelligence quotients were below 100 made lower accomplishment quotients. These six exceptions were all pupils whose intelligence quotients fell between 90 and 110.

Probably we cannot conclude from this that the bright pupils are practically all indolent, and that the dull pupils are all doing as well as or better than could be expected. Of course, in the extreme cases this conclusion is doubtless warranted. For example, one boy with an I.Q. of 141 had an A.Q. of only 89.5, another with an I.Q. of 138.5 had an A.Q. of 90.7, and a girl with an I.Q. of 155 had an A.Q. of only 85.6. On the other hand, pupils with I.Q.'s of 66, 77, and 78 have A.Q.'s of 108, 110, and 113, respectively. It is perhaps significant that the difference between the intelligence and accomplishment quotients of the dull pupils is less than the corresponding difference for the bright pupils. Doubtless, therefore, the bright pupils tend more strongly to work below capacity than the dull pupils do to over-work.

The correlation between intelligence quotients and educational quotients was, as would be expected, very high, notwithstanding the fact that a group test of intelligence was used to determine the intelligence quotients instead of the Stanford Revision of the Binet test, and notwithstanding the further fact that only one form of the group test was given. The coefficient of correlation, worked out by the Spearman footrule with Pearson's suggested correction, was found to be 0.90.

The correlation between the intelligence quotients and accomplishment quotients was a significant negative correlation, —0.46. The correlation between educational quotients and accomplishment quotients was —0.27.

It would appear, then, that the best basis for the classification of pupils in the intermediate and upper grades would be such a series of educational tests as the Stanford Achievement Tests. The pupils should be classified, promoted, or demoted on the basis of the educational age, while the educational quotient should be used as a basis for grouping the pupils of a grade into sections. The accomplishment quotient is valuable for diagnostic purposes. In the primary grades the individual Binet Tests should be the basis for classification, but the teacher's judgment should be considered in making these adjustments, so that such factors as health, personal traits, and environmental influences can be taken into consideration.

CHILD ACCOUNTING

I—PRESENT CONDITIONS

ARTHUR B. MOEHLMAN
University of Michigan and Detroit Public Schools

In recent years the attention of teachers and administrators has been brought to bear increasingly upon the need for adequate child educational records. The development of the scientific movement, the results of the surveys, and continuous individual investigations have shown the weaknesses and inadequacies of present local and state systems of records. The measurement movement has made great progress in the development of a scientific technique for the consideration of instructional data, but scarcely any general advance has taken place in the field of child-accounting records. Investigators have from time to time pointed to the need for uniformity and reliability but the importance of the general problem has received little recognition.

For purposes of this discussion the problem of child accounting will be defined as the recording of all activities, instructional and executive, that are necessary in the keeping of the essential records of the individual child during his school life. There are ultimately two complementary problems, pre-school records and post-school records, but only the phase relating to school records will be considered here.

Child accounting will be considered in three brief articles: (1) from the standpoint of existing conditions; (2) the historical aspects; and (3) the need for organization and uniformity.

What are some of the present conditions that make better child accounting necessary?

Let us start with the children themselves. It may be assumed for purposes of discussion that the compulsory school age is 7 to 16 years. There are many variations by states in the compulsory education laws but generally the 7 to 16 year age limits set the problem. If there were the proper accounting for these children and consequently rigid enforcement of the compulsory education

law, it would be assumed that practically all[1] such children would be in some type of school.

What are the actual conditions?

According to the 1920 federal census there were 20,683,361 children between the ages of 7 and 16 years and of this number 2,688,521 were not in school. In other words 13 percent of the children of compulsory school age were not recorded as receiving any educational service from the community. This does not mean that these children were not in school at any time during their childhood. The data are presented by age groups in Table I and Figure 1. The number not in school at age seven

TABLE I. SCHOOL ATTENDANCE AND NON-ATTENDANCE (UNITED STATES) ACCORDING TO THE 1920 CENSUS

Age	Total Census	Number in School	Number not in School	Percent not in School
5	2,347,839	441,411	1,906,428	81.2
6	2,338,315	1,480,714	857,601	36.7
7	2,287,609	1,905,404	382,205	16.7
8	2,273,280	2,010,894	262,386	11.5
9	2,151,032	1,944,314	206,718	9.6
10	2,233,704	2,077,965	155,739	6.9
11	2,097,782	1,970,255	127,527	6.1
12	2,234,759	2,082,749	152,010	6.8
13	2,028,627	1,877,429	151,198	7.5
14	2,046,265	1,766,784	279,481	13.7
15	1,861,445	1,357,345	504,100	27.1
16	1,972,958	1,001,701	971,257	49.2
17	1,855,191	642,360	1,212,831	65.4
18	1,910,046	413,619	1,496,427	78.3
19	1,830,934	252,680	1,578,254	86.2
20	1,781,102	148,352	1,632,750	91.7
21 and 60, older	886,520	344,789	60,541,731	99.4
Totals	94,137,408	21,718,765	72,418,643	76.9

is 16.7 percent; at age eight, 11.5 percent; and at age nine, 9.6 percent. At ages ten, eleven, and twelve the number not served drops to slightly less than 7 percent, increasing again with age thirteen until at age sixteen 49.2 percent are not in school.

[1] Exceptions are for employment and for non-attendance permits based on permanent illness and disqualifying physical or mental characteristics.

Apparently—for these are mass statistics and only capable of general analysis—the enforcement of compulsory attendance laws is most effective during ages nine to fourteen when the size of the child, out of school and not working, apparently makes him easier to trace. From ages seven to ten, and from fourteen to sixteen, inclusive, the inadequacy of reliable accounting makes it difficult to secure attendance.

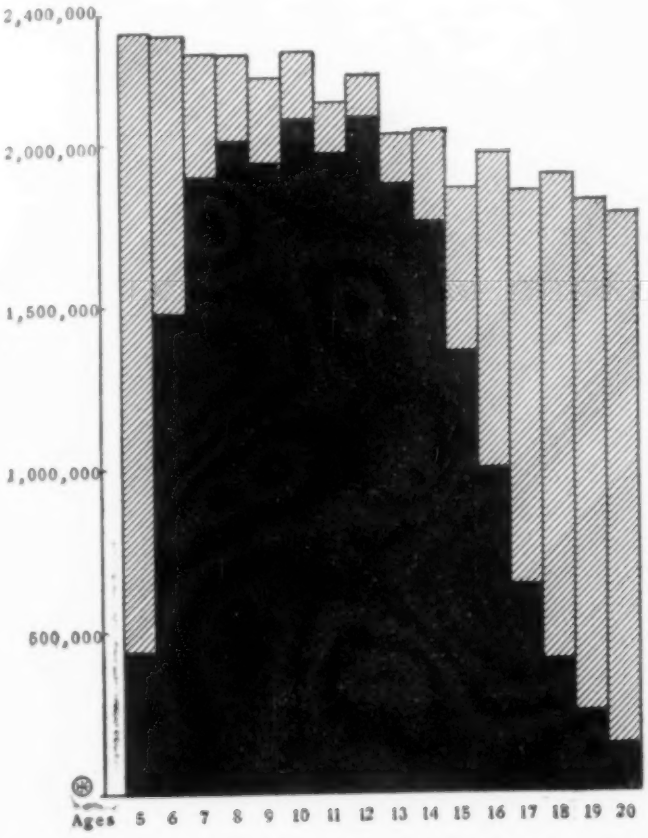


FIGURE 1. ATTENDANCE AND NON-ATTENDANCE OF SCHOOL CHILDREN IN THE UNITED STATES (CENSUS OF 1920).

(The black bars indicate the number in school, the shaded bars the number not in school. Data from Table 1.)

The significant fact that stands out boldly against the educational horizon is that 2,688,521 children in the United States in 1920 were receiving little if any service from the public or non-public schools. No attempt is made to consider the inadequacy of the service that is rendered through poorly trained teachers and poorly managed schools.

Possibly there may be some doubt in respect to the attendance statistics for the nation as a whole. In this general mass are included good, bad, and indifferent types of child accounting and law enforcement. It is probable that the sections where there is practically no enforcement of compulsory attendance present conditions so bad that they outweigh conditions in the more progressive communities, particularly the large cities.

What, then, are the actual conditions in a large progressive community? The city of Detroit may be taken as an example. Here child accounting has received considerable attention, and serious attempts have been made in the last five years to solve the problem. According to the last annual report of the superintendent of schools,[2] the situation is carefully analyzed.

In Figure 2 is shown the total number of Detroit children by age groups according to the 1922 school census. The black portion gives the number of each group in the public schools; the shaded area represents children in non-public schools; and the unshaded portion those not in school.

The total number of children served by community and private effort in June, 1922, was 186,726, or 77.3 percent of the total. Public schools had 137,201, or 56.8 percent of these in membership. According to the census report there were 11,730 children (7.1 percent) of compulsory school age not in attendance. A certain number of these can be accounted for by employment (1,532) and non-attendance (81) permits.

In comparison with the country as a whole this is a good showing. The 1920 federal census reports a total of 13 percent of children of school age who were not in attendance. Detroit,

[2] Detroit. *Eightieth Annual Report of the Detroit Public Schools (1922-23)* Detroit, Board of Education, 1923. pp. 5-7.

as would be expected for a large city, shows a larger proportion of children in school than does the country as a whole.

Actually, however, there were 7.1 percent (11,730) of the children of compulsory school age who were not in school. In

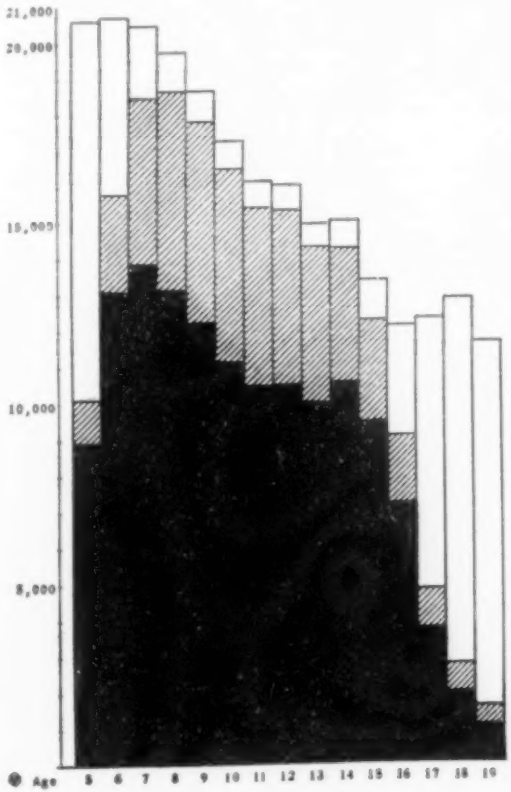


FIGURE 2. ATTENDANCE AND NON-ATTENDANCE OF SCHOOL CHILDREN IN DETROIT

(Black bars indicate attendance in public schools and shaded bars attendance in non-public schools. White bars indicate children not in school.)

terms of percents this is a relatively good condition, but when considered from the standpoint of the number of individual children a different side of the problem is presented.

After comparing Detroit conditions with national conditions,

TABLE II. RELATION OF SCHOOL CENSUS TO CHILDREN SERVED (DETROIT)

Age Group	Census 1922	Public Schools	Non-Public Schools	Total Served	Total Number Not in School	Percent not in School
5.....	20,613	8,920	1,194	10,114	10,449	50.9
6.....	20,712	13,100	2,666	15,766	4,946	23.9
7.....	20,485	13,855	4,596	18,451	2,034	9.9
8.....	19,727	13,128	5,525	18,653	1,074	5.4
9.....	18,629	12,218	5,555	17,773	856	4.6
10.....	17,224	11,137	5,323	16,460	764	4.4
11.....	16,095	10,449	4,917	15,366	729	4.5
12.....	16,007	10,490	4,825	15,315	692	4.3
13.....	14,916	10,009	4,258	14,267	649	4.4
14.....	15,004	10,537	3,703	14,240	764	5.1
15.....	13,341	9,458	2,768	12,226	1,115	8.3
16.....	12,096	7,191	1,852	9,043	3,053	25.2
17.....	12,287	3,719	1,068	4,787	7,500	61.0
18.....	12,839	1,959	757	2,716	10,123	78.8
19.....	11,624	1,031	518	1,549	10,075	86.7
Total...	241,599[a]	137,201	49,525	186,726	54,873	22.7

[a] Includes proportional distribution of 6,573 children not tabulated in original age-group calculations.

Superintendent Frank Cody comments upon the situation as follows:

> The apparent conclusion is that conditions in Detroit are extremely good as compared with the past and with the country as a whole, but a distinct and important problem is indicated that calls for a more scientific child accounting practice and a serious study of non-attendance in terms of social outcomes.

The second question to be considered is how are child records kept for those who are actually in school?

Except in a few cities, attendance records are kept on the basis of the Chicago rules[3] or modifications thereof. The result is a high percent of attendance that is without meaning. Under this system of accounting it is possible to show attendance as high as 97 percent.

How does it work?

[3] Superintendent Wells, 1862, provided that, in order to secure a good percent of attendance, a child should be "temporarily dropped" after five days of absence. These rules were generally adopted. Some cities have modified this to three days of absence.

After three days of absence a child is marked "left." The membership is automatically reduced and the percent of attendance—ratio of attendance to membership—is automatically increased. Investigation has shown that where emphasis is placed upon attendance the careful teacher will frequently ascertain whether an absent child is to be out more than three days and mark him left at once. Study of centers operating under such rules has shown that during the terrible influenza epidemic of 1918 the actual percent of attendance was higher than during normal times. Further study of these reports over a number of years compared with accounts kept without the "temporary left" expedient indicates that attendance percents vary from actual conditions by approximately 10 percent. In other words a 96 percent of average attendance is in reality only about 86 percent.

If all reports were kept on the same basis, some one may inquire, what would be the danger?

The grave danger in such procedure is that the inaccurate statement of facts leads to a misconception of conditions. It is fairly well established that in the northern parts of the country it is hardly possible to secure an average attendance, even under excellent conditions, of much better than 90 percent. The average over a period of years will probably be closer to 88 percent. Data secured on any other basis than strict accuracy merely serve to hide conditions. When we acknowledge conditions as they actually exist a problem is immediately presented.

In one city[4] the study of actual attendance led to the discovery that truancy and incorrigibility play only a very minor part in school absence. Ninety-two and six-tenths percent of absence was due to personal and family illness or other physical causes such as accidents. The knowledge of this fact resulted in a complete reorganization of the course of study in health education.[5] It also resulted in a new attitude towards attendance. Teachers came to realize that the health of the child was superior

[4] Pearl, N. H. and Geer, Florence. *Causes of Absence—Detroit Studies.* In manuscript.

[5] Pearl, N. H. and McLain, Rolf C. *Course of Study in Health Education.* Detroit, Board of Education, 1922, p. 48.

to classroom presence. Since the percent of attendance was no longer considered the index of school efficiency teachers sent children home who showed symptoms of illness instead of keeping them in surroundings where their possible infection might affect the entire group. Isolation of incipient infection will tend to maintain better general health conditions and, consequently, better general attendance.

The keeping of attendance percents in the past has apparently been for record purposes only. It would be difficult to find many cities where anything but *average* attendance is kept.[6] Average attendance is interesting from a statistical standpoint but of little practical value to the administrator or research man. The average figure for a semester or yearly period brings the high and low points in attendance into a meaningless whole. Even a study of monthly average attendance by grades shows that the variation by months is great and indicates a distinct administrative and health problem. A number of average attendance percents by months have been drawn in Figure 3. The significant fact is that the attendance starts to drop in October and continues until June. The use of the average apparently affects the warmer months of spring when attendance is known to be better than during the inclement winter time. The second noticeable factor is that as the children increase in age, attendance is uniformly better. Kindergarten attendance is poorest and that of the higher grades is best. If such data were kept on a monthly instead of a cumulative basis, another problem would be discovered that might lead to distinct changes in instructional policy and result finally in better school conditions. The importance of the attendance problem at present is familiar to anyone who has tried to develop the results of individual achievement through consecutive testing programs with the same group of children.

The third factor demanding study is the apparent lack of cumulative individual records. Conditions in this respect have been improved somewhat in larger centers since the introduction

[6] Over 100 cities canvassed had no record of anything but *average* attendance.

FIGURE 3. AVERAGE ATTENDANCE PERCENTS BY GRADES FOR DETROIT, 1922-23.

of the child accounting forms that grew out of the N. E. A. 1911-12 report. These records, however, carry only the academic records of the child. The problem is really broader than this. It demands child accounting that includes, in addition to academic data, health, social, emotional, vocational, achievement, and mental-test records as well.

When the larger cities are excepted, a study shows that there are few places where even a meagre cumulative record of academic achievement is carried from kindergarten through high school. Lack of unformity in marks and credits and absence of complete records makes difficult the transferring of children from one city to another, from city to rural district, and from rural district to city. The lack of cumulative records is partly responsible also for some of the retardation that stands out so vividly in age-grade studies.

The fourth factor to be considered is the grave need for constant appraisal studies of the child's progress through the school and the adjustment of the curriculum to the individual. These may be classified as the promotion-and-failure, the age-grade, and the age-grade-progress studies.

Promotion-and-failure studies may now be made from current records in most centers. The question arises as to how much of this work is done and what the results are. Important information in respect to the general adjustment of the curriculum to the children may be derived from such material. The average percent of promotions, however, is a meaningless figure. Variations among different schools and even among subjects within the same school are so great that careful surveys of this nature are necessary for any instruction department. In Table III are given the variations in high-school failure at Detroit for four semesters.[7] Note the variation by subjects by semesters and the great variation between subjects. Certainly these data are indicative of problems that must be carefully considered.

Thorndike and Ayres showed the possibilities and value of the age-grade study and for a time this device was widely used.

[7] Detroit. *Op. cit.*, pp. 110-111.

TABLE III. PERCENT OF HIGH-SCHOOL FAILURE BY SUBJECTS (DETROIT)

School Year	Semester	Languages	Exact Sciences	Social Sciences	Vocational	Health	Fine Arts	Total
1920–21	2nd	11.6	15.6	8.8	9.8	6.3	3.4	10.7
1921–22	1st	11.3	15.8	9.4	9.8	7.5	3.4	10.8
1921–22	2nd	12.5	16.4	10.0	9.6	8.6	5.0	11.7
1922–23	1st	14.6	18.1	11.0	11.0	7.9	5.0	13.1

Outside of the more progressive centers, however, the writer has been unable to discover that much is being done in a practical way with the age-grade survey. The age-grade study is a mass report and as such is limited in its possibilities; but it shows the problem of curricular adjustment and indicates the point of attack. Over a period of years it may be used as an index to determine progress in better classification.

The age-grade tables for one city (Detroit) are available since 1907 when the first survey was made. These data are shown in Table IV. They show the 1907 retardation to be 7.4 percent and the acceleration, 0.03 of one percent. In 1921 the retardation, as a result of better classification, was only 3.6 percent, and the acceleration 0.3 of one percent, although the membership had increased 240 percent. Certainly the labor and expense of making the age-grade study was justified on the basis of results as indicated by this table.

TABLE IV. RETARDATION AND ACCELERATION BY YEARS (DETROIT)

Year	Three Years or More Retarded		Three Years or More Accelerated	
	Number	Percent	Number	Percent
1907....	3,011	7.7	10	0.03
1908....	3,162	7.4	62	0.10
1913....	4,242	7.5	17	0.03
1914....	4,190	6.8	29	0.05
1915....	4,241	6.2	31	0.05
1916....	4,523	5.7	42	0.05
1917....	4,534	5.3	47	0.05
1918....	3,442	3.7	106	0.10
1919....	2,730	2.8	301	0.30
1920....	3,251	3.2	238	0.20
1921....	3,926	3.6	299	0.30

Even many of the larger cities find themselves handicapped in making an age-grade-progress study on account of inadequate records. In smaller districts this study is quite impossible. Yet its appraisal value is far greater than that of the age-grade study because it goes into the problem of individual adjustment much more thoroughly. It not only records the general retardation but shows individual progress through the grades in the retarded, the grade-at-age, and the accelerated groups.

A brief examination of conditions in child accounting has been made from the standpoint of (1) attendance, (2) attendance records, (3) individual records, and (4) appraisal records, and, from available evidence, it is apparent that a great field exists here that demands study, organization, and careful development.

THE PRIMARY CLASSIFICATION TEST

L. W. PRESSEY
Ohio State University

The examination to be described in the present brief article is essentially a revision of the writer's Primer Scale, issued four years ago. Under these circumstances, any detailed statement with regard to this new scale is evidently unnecessary.[1] The present revision, however, embodies features which in the writer's opinion are of general importance in any test work with young children, and the statistical methods used in evaluating the examinations are of general applicability.

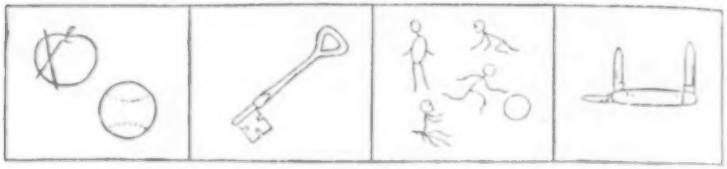
As a result of unusually varied and intensive experience in testing young children, the writer has become convinced that the fundamental problem in such work is that of administration. If the test is so planned that it is easily given and is interesting to the pupils, then nine-tenths of the difficulties usually met in such work will have been overcome—not only the obvious difficulties of administration, but the difficulties having to do with reliability and validity as well. It is the purpose of the present paper to describe the devices which have been employed to bring about such improved administration, and then to prove that a general increase in the value of the scale has been the outcome.

NATURE OF THE REVISION

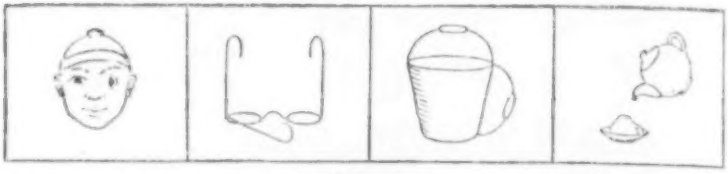
The revised scale, like the original Primer Scale, consists of four tests, each test presented on one page of a four-page folder. Each test has twenty items, thus making a total of eighty items or problems in the entire examination. In addition, the last

[1] The reader is referred to "A Group Scale of Intelligence for Use in the First Three Grades," *Journal of Educational Psychology*, 10:297-308, September, 1919; "A Group Scale of Intelligence for Use in the First Three Grades, Its Reliability and Validity," *Journal of Educational Research*, 1:285-94, April, 1920; also "Measuring the 'Usefulness' of Tests in Solving School Problems," *School and Society*, 12:531-534, November 27, 1920; and "The Influence of (a) Inadequate Schooling and (b) Poor Environment upon Results with Tests of Intelligence," *Journal of Applied Psychology*, 4:91-96, March, 1920.

three tests are each preceded by four items which are used as examples. The first four items of each test are shown in Figure 1.

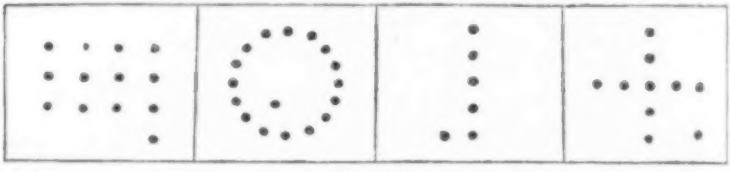


TEST 1



TEST 2

TEST 3



TEST 4

FIGURE 1. ITEMS FROM THE PRIMARY CLASSIFICATION TEST

In the first test the children are told to "put a cross on the part of the key that goes into the lock," to "cross out the biggest blade of the knife," and so, in like fashion, for the remaining sixteen items. In the second test the children are to "cross out the part that's wrong." In the third test the directions are to

"cross out the thing that doesn't belong with the others." In the last test the problem is to "cross out the dot that spoils the pattern."

The distinctive feature of the revision appears in the first test. During this test the children are thoroughly drilled in, and habituated to, the general procedure and method of the examination. The items are taken up by the examiner one by one; and, since each item is a picture and is referred to explicitly and unmistakably in the directions for that item, there is a minimal possibility that any child will lose his place on the page. The children thus learn to take up the items in order, going across the page from left to right and working down the page from the top row to the bottom. The children are also told, on each of the twenty items of the first test, to "cross out" or "make a cross." In this way the pupils become habituated to the form of response, which is used throughout the examination, and the first test thus serves both as a real test and as an extended practice exercise.

The four tests of the revision are carefully arranged in order with reference to ease in giving. More pictures have been used than appeared in the original Primer Scale, and they are for the most part pictures of objects which are of interest to young children.[2] The examination has been shortened to eliminate any possibility of fatigue, and the "typography" has been much improved. The sheet of teacher's directions has also been made clearer and more complete. It includes record sheets, norms, and interpretative data.

It will be obvious to those familiar with the original Primer Scale that the essential nature of that test has hardly been changed in the revision. Its valuable features have been retained, namely, the question-and-answer method of giving the directions, the cross-out method of indicating answers, and the ease and objectivity of scoring.[3] The changes have almost all been made

[2] The form-board test of the original "Primer" was dropped, partly because it was uninteresting and partly because it was very difficult to explain to first-grade children.

[3] For a discussion of these features the reader is referred to the first article mentioned in the first footnote of this article.

with reference to the improved administration of the tests. Now, what has been the result of these modifications?

The Value of the Scale for Use with First- and Second-Grade Children

In evaluating the Primary Classification Test[4] the first effort has been to determine the relation of scores on this group test to the results on the Revision of the Stanford-Binet Scale. In this work the writer was so fortunate as to secure the splendid cooperation of Superintendent B. O. Skinner of Marietta, Ohio. All children who enter the schools of Marietta are given the Stanford-Binet test. In the spring of 1922 all children in the first two grades were also given the Primary Classification Test—making a total of 424 cases who had taken both tests. The remainder of the paper deals with these results.

The first question considered was whether scores on the Primary Classification Test were in accord with Stanford-Binet "norms;" recent investigations have demonstrated that the norms on certain group tests are astonishingly incomparable.[5] To obtain an answer to this question the Classification-Test scores of the children having Binet mental ages of six, seven, and eight were tabulated, and the median Classification-Test scores were found for each Binet mental age. For the three mental ages mentioned these medians were found to be 36.5, 46.5, and 53.5. The Primary-Classification-Test norms for these three ages are 35, 47, and 55. Evidently standards for the two examinations are largely similar; mental ages obtained from the Primary Scale will average with Binet mental ages.

This is good as far as it goes; it shows that the children on whom the Primary Scale was standardized were of about the same average intelligence as the children tested by Terman in making the Stanford Revision. But the important practical

[4] The Primary Classification Test is not intended for use in the third grade, except with III-B classes somewhat backward in reading; the old Primer Scale was not found to be highly satisfactory in the third grade.

[5] Stenquist, J. L. "Unreliability of Individual Scores in Mental Measurements." *Journal of Educational Research*, 4:347-54, December, 1921.

question still remains as to whether there is close agreement between the two scales, in individual cases. To determine this matter, mental ages on the Stanford Scale[6] were first correlated with Primary-Classification-Test scores. The correlations were:

	Grade						Age		
	I-B	I-A	II-B	II-A	I	II	6	7	8
Correlation	.65	.74	.69	.64	.69	.71	.69	.70	.83

For the entire group the correlation was 0.87. These correlations may be compared with analogous correlations obtained for the original Primer Scale. (See second reference in footnote 1.) The correlation found for six-year-old children between Binet Mental Age and Primer Scale was 0.62, the correlation between Binet and Primer score for a group covering the first three grades was 0.66. There is surely here a suggestion of a definite improvement in the Primary Classification Test over the original Primer Scale.[7]

Correlation coefficients do not show, however, the information which the practical school man most needs; they do not indicate the extent to which confidence may be placed in the score of a given child. To obtain some information on this matter, the scores on the Primary test were converted into mental ages,[8] and the differences found between them and Binet mental ages. Thus, suppose a child showed a Primary mental age of 8 years

[6] The Binet test had been given at entrance to school; thus the II-A children had been tested a year and a half ago, the II-B children a year ago, and so on. It was thus necessary to calculate, from the intelligence quotient, Binet mental ages at the time of the Primer-Scale testing. So any factors causing variations in the intelligence quotient will lower these correlations. In other words, the dice are weighted against the Primary Scale.

[7] The correlation between Primary Classification Intelligence Quotient and Binet Intelligence Quotient was found to be 0.79. Correlations with intelligence quotients were not worked out for the original Primer test. But comparisons with some results reported by Guiler ("How Different Tests Agree in Rating Children," *Elementary School Journal*, 22:734-44, June, 1922) is of some interest. Guiler obtains correlations, for a group of 63 children in the sixth, seventh, and eighth grades, of 0.88 between the Stanford-Binet and the Illinois Intelligence Test and 0.75 between Binet and combined Scale A and Scale B of the National tests.

[8] This was done by making a graph from the Primary-Classification age norms and reading the mental ages from this graph.

6 months and a Binet mental age of 8 years 11 months. The difference in mental status as indicated by the two tests would then be five months. Such differences were found for all the children, and the median difference was then determined.[9] This median difference between the scales, in the mental ages assigned to the children, was found to be eight months. The median difference in intelligence quotients was 7.9 points.

It is impossible to compare the Primary Classification with the original Primer Scale as regards these last figures, since data for the original scale were not so handled. But similar data are available, regarding the relation between Binet results and results on other group tests. Thus Garrison and Tippett[10] found, for a group of children averaging twelve years of age chronologically, a median difference in mental ages of 17 months between the Binet and the Otis scales. Otis and Knollin[11] report a median difference between two halves of the Binet of 10.6 months.

More comparable are the findings of a median difference in intelligence quotients of 7 points between the Binet and the National Intelligence Tests, Forms A and B combined, and of 9 points between the Binet and the Illinois Intelligence examination for 63 children in the sixth, seventh, and eighth grades. These findings are said to be more comparable because they are expressed in terms of the intelligence quotient and so are largely

[9] The concept is essentially the concept of the probable error, as developed by Otis for measuring the reliability or agreement of a test with itself, and for measuring the validity or agreement of a test with some independent criterion. And this application has all the advantages of Otis' methods—independence of heterogeneity of the group, significance with reference to the individual case, and so on. In fact, for a large proportion of the situations in which an indication regarding relationship is desired in test work, some such simple method would seem rather more valuable than the abstruse procedures of correlation.

[10] Garrison, S. C., and Tippett, J. S. "Comparison of the Binet-Simon and Other Tests," *Journal of Educational Research*, 6:42-48, June, 1922. It should be pointed out that some of the differences reported in the papers referred to here are probably due to differences in norms. But Stenquist, after carefully adjusting his norms with reference to this matter, found a median difference between Otis and Haggerty test mental ages of nine months, and between Otis and National Intelligence Test, Form A, of 13 months. (See reference footnote 5.)

[11] Otis, A. S., and Knollin, H. E. "Reliability of the Binet Scale and of Pedagogical Scales," *Journal of Educational Research*, 4:121-43, September, 1921.

independent of differences in the range of scores from one chronological age to another. Lastly may be mentioned, from one of the numerous studies of the constancy of the intelligence quotient, the findings of Miss Martens, who reports a median difference of five points between successive applications of the Binet Scale.[12] The Primary Classification Test surely stands up fairly well in these comparisons.

It is to be emphasized, however, that even the median differences mentioned above do not give a school official that fullness of information which he should have, for judicious use of test findings in dealing with individual cases. From the above statements a principal will understand that no more than half his children will, if tested by both Primary and Binet, differ more than eight months in mental age. The principal, however, desires to know not merely the general run of the differences, but the danger of being seriously mistaken, which he may incur in inferring mental age from the Primary Classification Test. Not the median difference, but the number of extreme differences, is the crucial problem for him. The writer would like to suggest the following simple table as a very useful common-sense presentation of findings for the needs of such a practical school man. This table shows the percent of the total number of cases in which differences between Binet and Primary mental ages will amount to 0 to 3 months, 4 to 7 months, etc. From this table a principal can see at once what chance there is that a pupil's mental age according to the Primary Scale will differ by an indicated amount from his mental age according to the Binet Scale.

	Difference in Mental age							
	0–3	4–7	8–11	12–15	16–19	20–23	24–27	28-or over
Percent of Cases.....	26	26	20	13	8	4	2	1

There is evidently only a 1 percent chance that a case will differ 28 months or more in mental age, if tested with the two scales, a 3 percent chance (1 plus 2) that a case will differ 24

[12] Martens, Elise H. *A Study of Individual Retests.* University of California, Bureau of Research Studies No. 7.

months or more, a 28 percent chance that the case will differ more than a year, and a 26 percent chance that the mental ages will not differ more than three months. The principal has at least—as the writer sees it—that information which he needs, in dealing with a child on the basis of the group test.

Analogous handling of the intelligence quotient differences, including Guiler's data, is shown in the table below:

	Differences in I. Q.							
	0–4	5–9	10–14	15–19	20–24	25–29	30–34	35 and above
Percent of Cases								
Binet and Primary....	33	27	19	13	4	2	1	1
Binet and National...	129	32	21	8	8	1	1	0
Binet and Illinois.....	25	32	18	14	6	5	0	0

Discussion

It was stated at the beginning of the paper that two points would be dealt with (1) improvements in the Primary Classification Test over the original Primer Scale leading to greater efficiency in the giving of the test and (2) evidence that these improvements in the technique of administering had resulted in a scale giving more accurate measurements than the original Primer Scale—resulting, in fact, in a group test of intelligence for the first two grades which compares as favorably with the Stanford-Binet as do certain well-known group scales for the upper grades. The writer feels that the evidence presented is reasonably adequate on these two points, but the last tables have raised a further question.

Very likely the reader has been astonished at the disagreement which may be found occasionally between two tests even when the correlations are as high as 0.70 or 0.80. The writer feels it a disagreeable duty to emphasize that this may be so; to emphasize that a high correlation is entirely compatible with marked variations between the two measures studied; and to point out that some such method of statement is needed as the percentage tables above if the facts are to be presented as they should be for use in connection with individual diagnosis. But

the question in the mind of the reader is doubtless as to whether, under the circumstances, the Primary Scale can be considered of any dependability in dealing with individual children.

It is to be mentioned, in the first place, that the evaluation of the group tests has been throughout in terms of the Binet, but that the Binet is a much more fallible instrument, itself, than most people realize. At the meetings of the American Psychological Association, in 1916, Doll startled his audience by evidence that half of the Binet scale was more effective than the whole scale in distinguishing the feeble-minded, and Brigham presented data to show that for certain ages certain tests in the Binet were passed more readily by defective than by normal children.[13] It is a deplorable fact that since this time few studies of this problem have been made. Investigators seem to be more anxious to develop applications of the Binet than to find out whether it is worth while! But at least the two studies above mentioned stand uncontradicted.[14] There is, then, warrant for suggesting that when the Binet test differs from some other test, the Binet and not the other test may occasionally be at fault. The writer believes it reasonable to conclude that the Primary Test, if used with due regard to other sources of information, may be considered an instrument of some considerable value in the study of first- and second-grade children—at least about as valuable as certain group tests which are widely used in the upper grades.

It seems evident, however, that in the diagnosis of individual cases group tests in general do not give results as trustworthy, as is desirable. The question is whether the trustworthiness of these scales can be increased. The writer is inclined to believe

[13] Doll, E. A. "A Brief Binet-Simon Scale," *Psychological Clinic*, 11: 197-261. December, 1917. Brigham, Charles C. *Two Studies in Mental Tests.* Princeton, New Jersey, Psychological Review Company, 1917. 254 pp. (The Psychological Monographs. Vol. 24, No. 1, Whole No. 102.)

[14] Rather, the findings have been essentially confirmed. See for instance Burt, Cyril, *Mental and Scholastic Tests*, London, London County Council, 1921. pp. 199-206. It may be added that Burt presents partial correlation analyses (using criteria, to be sure, which seem to the American investigator rather crude) to indicate that scores on the Binet are more influenced by schooling than by native ability.

that check or "verifying" tests may thus function, but space does not permit the discussion of this procedure at this time.

Summary

1. The paper presents a revision of the Pressey Primer Scale, and this revision has certain marked improvements especially as regards the method of giving it.

2. Results are presented to show the general agreement of this test with the Stanford-Binet as regards norms, and the correlation of scores on these two scales.

3. Data are presented, making use of certain new methods, to indicate the value of the scale in diagnosing the individual child.

4. A suggestion is made regarding improvement in the reliability of group tests of intelligence.

THE ACCURACY OF CERTAIN STANDARD TESTS FOR SCHOOL CLASSIFICATION

PERCIVAL M. SYMONDS
University of Hawaii

We have passed through a period of enthusiastic acceptance of the use of mental tests for the purposes of school classification. Quite widely the tests have been hailed as scientific instruments whose use would make school grading more precise. The leaders of the testing movement have been quite well aware of the inaccuracy of the tests and in many cases have precisely defined it. But their statements have been in terms rather far removed from the practical necessities of school administration. The school principal has not been able to interpret these statements—which have been in terms of "coefficients of correlation" or "probable errors"—to fit his own problems. The purpose of this paper is to state the accuracy of certain standard tests in the customary terms of the statistician and then to interpret these statements for their practical application in school.

The question of the accuracy of mental tests resolves itself into three factors. The first factor is technically called "reliability." By this the psychologist means the degree to which the scores of a second giving of a test correlate with the scores of the first giving. In the actual determination of reliability two methods are in vogue. One is to divide the test into halves, chosen at random, and to correlate the scores of these halves against each other. Then a correction by the Spearman (Brown) formula is made to account for the fact that the whole test is twice as long as each of the two halves. The other method is to correlate one test with a second test as nearly as possible like the first in form, construction, and difficulty but different as to material—for example, Forms 1 and 2 of Scale A of the National Intelligence Test. The method of correlating random halves gives higher reliabilities than the method of correlating two forms of the same test because it does not contain those differ-

ences coming from a new day, a new room, and a new examiner, which a second giving may produce. But it is precisely these factors which the school administrator holds partly responsible for the inaccuracy of his tests. Consequently the method employed in this paper for finding reliabilities is to determine the correlation between the first and second forms of the same test, given on different days by different examiners. The tests were:

National Intelligence Test, Scale A, Forms 1 and 2 (N.I.T.,A)
National Intelligence Test, Scale B, Forms 1 and 2 (N.I.T.,B)
Thorndike-McCall Reading Scale, Forms 1 and 2 (Th-McC)
Woody-McCall Mixed Fundamentals, Forms 1 and 2 (Wσ-McC)
Two spelling scales of 50 words each, each scale containing different words. Ten words were selected from each of the five columns O, Q, S, U, W, of the Buckingham Extension of the Ayres Scale.
Two compositions scored by the Nassau County Supplement of the Hillegas Scale. Subjects: "What I Should Like to do Next Saturday" and "An Exciting Experience."

The testing was done in grades IV to VIII inclusive of the Territorial Normal Training School. The N.I.T. Scale A, Form 1 and the first forms of all the subject-matter tests were given on the morning of November 21, 1922. The N.I.T. Scale B, Form 1 was given on November 22. The N.I.T. Scale A, Form 2 and the second forms of the subject-matter tests were given on November 27. The N.I.T. Scale B, Form 2 was given on November 28.

To determine the reliability, I used the formula derived by Otis ('21),

$$\text{P.E. } (\textit{single measure}) = \frac{0.6745}{\sqrt{2}}\sigma_{(A-B)} = 0.4769\ \sigma_{(A-B)}$$

where A—B is the difference between scores on the two forms of the test for any individual. For the tests used in this study, I have given in Table I these probable errors (calling them $\text{P.E.}_{\text{score}}$). I have also given, under the heading $\text{P.E.}_{\text{grade}}$, the part of a yearly grade within which 50 percent of the pupils are accurately placed by each test.

TABLE I. PROBABLE ERRORS OF CERTAIN STANDARDIZED TESTS

Test	Number of Cases	P. E. score	P. E. grade
N. I. T., Scale A			
Test 1	232	1.7	0.8
Test 2	232	2.3	0.6
Test 3	232	2.4	0.7
Test 4	232	3.0	0.7
Test 5	232	2.6	0.8
Total	232	5.4	0.3
N. I. T., Scale B			
Test 1	242	1.8	0.5
Test 2	243	1.7	0.6
Test 3	243	2.7	0.9
Test 4	243	1.7	0.8
Test 5	243	2.5	0.8
Total	242	4.9	0.3
Composite, Scales A and B	200	7.8	0.2
Thorndike-McCall	232	2.6	0.5
Woody-McCall	229	1.4	0.5
Spelling	226	2.0	0.2
Composition	237	0.42	0.7

In passing it may be noted that the same result is obtained for the P.E. of a score if we use the formula

$$\text{P.E.}_{\text{score}} = 0.6745\sigma_{\text{distribution}}\sqrt{1-r_{AB}}$$

where A and B are the scores on the two forms of the test given.

Besides these probable errors, the probable errors of composites of the above tests are of interest. To obtain the probable error of the composite of two tests, each of whose probable errors is known, the formula for the probable error of the sum of two variables may be used.

The formula is $\text{P.E.}_{x+y} = \sqrt{\text{P.E.}_x^2 + \text{P.E.}_y^2}$, unless there is correlation between x and y. In that event a third term, namely $2r\,\text{P.E.}_x\text{P.E.}_y$, must be introduced under the radical. In this investigation x is the difference in scores between two forms of a given test, and y has the same meaning for another test. These differences are apparently practically uncorrelated. For example, differences between scores on Forms 1 and 2 of the National Intelligence Test, Scale A and differences between scores in Forms 1 and 2 of the Thorndike-McCall Reading Test were found to be correlated to the extent of -0.07; while Thorndike-

McCall differences were related to Woody-McCall differences to the extent of + 0.03.

These two correlations are of special significance. They indicate that the reliability of a test depends not at all on the *general* condition of the individuals taking the test—state of health, fatigue, or nervousness, but on factors which are peculiar to each test in itself. We may confidently throw out of court all those criticisms of test reliability which depend on the general condition of the individual and may look for the causes of the unreliability of a test in the conditions of the individual unique for that test, in the nature of that test itself, and in the special conditions under which it is given.

By means of the formula for the probable error of the sum of variables, the probable errors for composites of the tests were found as shown in Table II.

TABLE II. PROBABLE ERRORS OF COMPOSITES OF CERTAIN STANDARDIZED TESTS

Composite	P. E. score	P. E. grade
N. I. T. (A)+Th.-McC.	6.0	0.3
N. I. T. (A)+Th.-McC.+Wo.-McC.	6.2	0.3
N. I. T. (A)+Th.-McC.+Wo.-McC.+10 Composition	7.5	0.3
N. I. T. (A)+Th.-McC.+Wo.-McC.+10 Composition+Spelling	7.7	0.2
N. I. T. (A+B)+Th.-McC.	7.8	0.2
N. I. T. (A+B)+Th.-McC.+Wo.-McC.	8.0	0.2
N. I. T. (A+B)+Th.-McC.+Wo.-McC.+10 Composition	8.9	0.2
N. I. T. (A+B)+Th.-McC.+Wo.-McC.+10 Composition+Spelling	9.2	0.2
Th.-McC.+Wo.-McC.	3.0	0.4

To the results shown in Table II, I add those of Table III in which the following weights are used in making composites: National Intelligence Scale, 1; Thorndike-McCall, 4; Woody-McCall, 3; Composition, 10; and Spelling, ½. These weights were selected first to equalize variability and secondly to weight the tests roughly in the ratio intelligence 5, reading 5, arithmetic 4, composition 3, spelling 1. These latter ratios are the result of a consensus of opinion in a class in educational measure-

ments regarding the importance of school subjects as criteria for classifying in the intermediate grades. Regression methods with suitable criteria would give more exact weights, but probably the results would not differ radically from those obtained by using the weights above suggested.

TABLE III. PROBABLE ERRORS OF WEIGHTED COMPOSITES OF CERTAIN STANDARDIZED TESTS

Composite	P.E. score	P.E. grade
N. I. T. A+4 Th.-McC.	11.9	0.3
N. I. T. A+4 Th.-McC.+3 Wo.-McC.	12.6	0.3
N. I. T. A+4 Th.-McC.+3 Wo.-McC.+10 Composition	13.3	0.3
N. I. T. A+4 Th.-McC.+3 Wo.-McC.+10 Composition+½Spelling	13.3	0.2
N. I. T. (A+B)+4 Th.-McC	12.9	0.2
N. I. T. (A+B)+4 Th.-McC.+3 Wo.-McC.	13.5	0.2
N. I. T. (A+B)+4Th.-McC.+3 Wo.-McC.+ 10 Composition	14.1	0.2
N. I. T. (A+B)+4Th.-McC.+3Wo.-McC.+10 Composition+½Spelling	14.2	0.2

If the P.E.score is computed for each class, it is found to differ for different rooms and grades. For example, for N. I. T. (A) it ranges from 3.4 to 6.9 and for Thorndike-McCall from 1.3 to 3.3. Thus, a given test shows different reliability in different classes. Part of this is due to mere chance, and is a result of the small number in the class, but part of it is also due to actual differences in variability. Conditions in the class as a whole as well as in the individuals and in the test itself account for the unreliability of a test. The P.E. for each grade in the N. I. T. correlates with the P.E. for each grade in the Thorndike-McCall to the extent of 0.3. But the P.E. of the N. I. T. (A) for each room correlates with the average N. I. T. (A) scores for that room to the extent of 0.6, the P.E. of the Th.-McC. for each room correlates 0.7 with the average Th.-McC. scores for that room, while the N.I.T. (A) scores correlate with the Th.-McC. scores 1.0 according to the rank formula. Making the scores constant by partial correlation, the P.E. of the N.I.T. correlates practically zero with the P.E. of the Thorndike-McCall.

This would indicate that the differences in unreliability in various classes can be partly explained by the tendency of the unreliability in terms of P.E. to grow larger as the level of the class in gross scores grows higher, but that there is no general tendency of a class to have a low or high reliability. The causes for variation in reliability in a class must be sought in the special conditions under which the test is given and not in some general variability of the class.

The correlation between the size of the score and the size of the difference between the first and second test (irrespective of sign) is + 0.12 for the N.I.T. Scale A, 0.00 for the N.I.T. Scale B, + 0.16 for the Thorndike-McCall, and + 0.23 for the Woody-McCall, indicating a small tendency for the P.E. of a score to become large with the score itself. The regression equation for the $P.E._{score}$ from a given gross score on the N.I.T. Scale A is $Y = .061X + .78$. Tabulating certain values gives:

Score	$P.E._{score}$
30	2.6
60	4.5
90	6.3
120	8.1

From this it appears that the National Intelligence Test is more accurate in the placing of a child in the lower grades than in the upper grades. Whereas the difference in score from one grade to another is greater in the lower grades, the P.E. of the scores is much less among the lower scores. The accuracy of the National Intelligence Scale A expressed as the part of the year within which it places 50 percent of the children of a given grade accurately is:

Grade	$P.E._{grade}$
III	0.1
IV	0.2
V	0.4
VI	0.5
VII	0.5
VIII	0.9

Even though the P.E. remained the same with increasing scores this fact would remain because of greater differences between standard scores from one grade to another in the lower

grades. This suggests that if the tests were to be revised, some of the easier material might be weeded out and some of the more difficult material might be filled in. Such a procedure would doubtless make the tests more generally useful for classification purposes in grades from the third to the eighth.

A warning should be given with regard to the two tables which appear in the preceding paragraph. They are only tentative and probably highly unreliable. The correlation of + 0.12 has a P.E. of 0.04. Two or three cases of low scores with a large gain in the second test would be quite sufficient to lower the correlation coefficient, and to change the regression equation and the tables considerably. It is undoubtedly true that the P.E. does increase with the score, but just how much cannot be stated on the basis of 232 cases. The above regression equation is of more interest to illustrate method than to give actual constants.

The school administrator would like to know how many times he must repeat a test before he obtains a certain accuracy of grading. A practical form of this "accuracy of grading" may be twofold. First, it may be required that a certain percent of the pupils be accurately classified; and, second, a degree of fineness of grading may be specified. For example, the requirement may be that 90 percent of the pupils be graded correctly to within half a year. If we let x denote the distance from the mean of a normal distribution which includes the given percent of cases, g the grade unit or difference between the norms of two consecutive grades,[1] and h the part of the year within which accuracy of classification is required, the formula for n, the number of necessary repetitions of the test is,

$$n = \left(\frac{\text{S. D.}_{\text{score}} \times x}{g \times h} \right)^2$$

For use in this formula, x may be read from a table of the fractional parts of the area of a normal curve corresponding to distances from the mean along the base line. Half the given per-

[1] In our calculations we used one-fourth of the difference between the norms of fourth and eighth grades.

cent must be taken, since the distance is counted both ways of the median.[2]

The tables below indicate for certain tests and for certain composites of tests the number of times they must be repeated before certain accuracies of grading are reached.

TABLE IV. NUMBER OF REPETITIONS OF THE NATIONAL INTELLIGENCE TEST FORM A NECESSARY TO GRADE STATED PERCENTS OF PUPILS WITH SPECIFIED ACCURACY

Percents of Pupils	Repetitions Necessary to Grade within Indicated Fraction of a Year			
	1.00	0.75	0.50	0.25
50	1 (0.1)	1 (0.2)	1 (0.4)	2 (1.3)
75	1 (0.3)	1 (0.6)	2 (1.2)	5 (5.0)
80	1 (0.4)	1 (0.7)	2 (1.5)	7 (6.2)
90	1 (0.6)	2 (1.1)	3 (2.6)	11 (10.2)
95	1 (0.9)	2 (1.6)	4 (3.6)	
98	2 (1.3)	3 (2.3)	6 (5.1)	
99	2 (1.6)	3 (2.8)	7 (6.2)	

Table IV reads as follows: To grade 50 percent of the pupils in the intermediate grades accurately to within one year, one giving of the test is sufficient (0.1 of a test to be exact). To grade 90 percent of the pupils accurately to within 0.50 of a year, three givings of the test are necessary (2.6 givings of the test to be exact).

[2] For such a table see Rugg, H. O. *Statistical Methods Applied to Education*. Boston, Houghton Mifflin Company, 1917. p. 389.

TABLE V. NUMBER OF REPETITIONS OF THE THORNDIKE-McCALL READING SCALE NECESSARY TO GRADE STATED PERCENTS OF PUPILS WITH SPECIFIED ACCURACY

Percents of Pupils	Repetitions Necessary to Grade within Indicated Fraction of a Year			
	1.00	0.75	0.50	0.25
50	1 0.3	1 0.5	2 1.1	5 4.7
75	1 0.9	2 1.5	4 3.4	14 13.6
80	2 1.1	2 1.9	5 4.2	
90	2 1.8	4 3.1	7 7.0	
95	3 2.5	5 4.4	10 9.9	
98	4 3.5	7 6.2		
99	5 4.4	8 7.6		

Of the three factors mentioned at the beginning of this paper as involved in the accuracy of tests, the first, the reliability of the tests, has now been considered. The second factor is the difference in difficulty between two forms of the test. Although this does not involve the accuracy of a single test, it must be considered when the test is to be repeated for accuracy of grading. The third factor involved in the accuracy of tests is the gain due to practice effect. This also does not involve the accuracy of a single test but must be considered when the results of two forms of the same test are combined. Thorndike has shown that this practice effect amounts to considerable, although its influence rapidly falls off after a third giving. It has been suggested that the reliability is increased after a test has undergone

TABLE VI. NUMBER OF REPETITIONS OF THE WOODY-MCCALL MIXED FUNDAMENTALS SCALE NECESSARY TO GRADE STATED PERCENTS OF PUPILS WITH SPECIFIED ACCURACY

Percents of Pupils	Repetitions Necessary to Grade within Indicated Fraction of a Year			
	1.00	0.75	0.50	0.25
50	1 0.2	1 0.4	1 0.9	4 3.8
75	1 0.7	2 1.2	3 2.7	11 11.0
80	1 0.8	2 1.5	4 3.4	14 13.5
90	2 1.4	3 2.5	6 5.6	
95	2 2.0	4 3.5	8 7.9	
98	3 2.8	5 4.9	12 11.1	
99	4 3.4	7 6.1		

several repetitions. As yet there have been no studies published to validate this, although the increase in correlations between functions after practice noted by H. L. Hollingworth, A. I. Gates, and others may be partly due to this increased reliability. If, however, the results of two forms of the same test are combined the gain of the second test over the first is so considerable that merely doubling the grade or age norms does not give valid standards for comparison.

I regret that I gave all my tests in the same order—the second form after the first—so that practice effect and difference in difficulty cannot be disentangled. Where the scores on the second form are lower than on the first form, one can safely say that the second form is harder than the first; but how much harder

TABLE VII. NUMBER OF REPETITIONS OF AN UNWEIGHTED COMPOSITE OF FORMS A AND B OF THE NATIONAL INTELLIGENCE SCALES NECESSARY TO GRADE STATED PERCENTS OF PUPILS WITH SPECIFIED ACCURACY

Percents of Pupils	Repetitions Necessary to Grade within Indicated Fraction of a Year			
	1.00	0.75	0.50	0.25
50	1 0.1	1 0.1	1 0.2	1 0.9
75	1 0.2	1 0.3	1 0.6	3 2.6
80	1 0.2	1 0.4	1 0.8	4 3.2
90	1 0.3	1 0.6	2 1.3	6 5.2
95	1 0.5	1 0.8	2 1.9	8 7.4
98	1 0.7	2 1.2	3 2.6	11 10.4
99	1 0.8	2 1.4	4 3.2	

cannot be stated with accuracy. This is true of Test 4 of Scale A, N.I.T., and Test 2 of Scale B. The second forms of these tests are more difficult than the first forms, but just how much more difficult cannot be stated with accuracy.

With regard to the Thorndike-McCall Reading Test, Gates finds "that seven days of practice produced no measurable improvement." I find that Form 2 gives scores 1.8 lower on the average than Form 1. Gates found that Form 2 gave scores 1.3 higher than Form 1. The discrepancy can be explained, I believe, in several peculiarities in the scoring of Form 2 which do not occur in Form 1. In Question 7 of Form 2 "in the woods" is a correct answer while "woods" is incorrect; in Question 24

TABLE VIII. NUMBER OF REPETITIONS OF AN UNWEIGHTED COMPOSITE OF THE THORNDIKE-MCCALL READING SCALE AND THE WOODY-MCCALL MIXED FUNDAMENTALS SCALE NECESSARY TO GRADE STATED PERCENTS OF PUPILS WITH SPECIFIED ACCURACY

Percents of Pupils	Repetitions Necessary to Grade within Indicated Fraction of a Year			
	1.00	0.75	0.50	0.25
50	1 0.2	1 0.3	1 0.6	3 2.4
75	1 0.4	1 0.8	2 1.8	8 7.1
80	1 0.6	1 1.0	3 2.2	9 8.8
90	1 0.9	2 1.6	4 3.6	
95	2 1.3	3 2.3	6 5.1	
98	2 1.8	4 3.2	8 7.3	
99	3 2.2	4 4.0	9 8.9	

"In Pawnee County" is a correct answer, while "Pawnee County" is not accepted; in Question 5 "he killed the fox" is a correct answer while "killed" is incorrect. For children of Oriental parentage an answer omitting the preposition is particularly easy and the "pidgin English" which is the language of the street here in Honolulu makes the children of this school liable to these errors. In scoring, the key was followed strictly. These are discrepancies that would not occur in a "controlled answer" test.

I find the gain of Form 2 over Form 1 of the various tests to be as shown in Table IX. As I stated above, this is both a matter of practice effect and difference in difficulty. The gain

of a second giving over the first for a composite may be found by simply totaling the gains for the separate tests.

The correlation between size of score and the difference in scores between first and second test is — 0.22 on the N.I.T., Scale A, — 0.09 on the N.I.T., Scale B, — 0.71 on the Thorndike-McCall, and — 0.14 on the Woody-McCall. Negative correlation is also reported by Thorndike ('22) and by Renshaw ('23). The regression equation for predicting the gain of Form 2 over Form 1 on the N. I. T., Scale A, given the original scores is

$$Y = 15.97 - 0.079X,$$

from which Table X is derived.

TABLE IX. GAINS OF FORM 2 OVER FORM 1

Test	Number of Cases	Difference (2−1)
N. I. T., Scale A		
Test 1	232	1.1
Test 2	232	2.0
Test 3	232	3.4
Test 4	232	−0.4
Test 5	232	4.0
Total	232	10.0
N. I. T., Scale B		
Test 1	242	0.9
Test 2	243	−1.2
Test 3	243	1.7
Test 4	243	2.4
Test 5	243	3.4
Total	242	7.0
Composite of Scales A and B	200	17.3
Thorndike-McCall Reading	232	−1.8
Woody-McCall Fixed Fundamentals	229	0.5
Spelling	226	1.3
Composition	238	−0.3

TABLE X. GAINS OF FORM 2 OVER FORM 1 FOR SCORES ON THE NATIONAL INTELLIGENCE TEST, SCALE A

Score	Gain of Form 2 over Form 1
30	13.6
60	11.2
90	8.9
120	6.5
150	4.1

The regression equation for predicting gain of Form 2 over Form 1 for N.I.T., Scale B, is

$$Y = 9.91 - 0.036X,$$

from which Table XI is derived.

TABLE XI. GAINS OF FORM 2 OVER FORM 1 FOR SCORES ON THE NATIONAL INTELLIGENCE TEST, SCALE B

Score	Gain of Form 2 over Form 1
30	8.8
60	7.8
90	6.7
120	5.6
150	4.5

The caution given above with regard to the correlations between score and P.E. should be repeated here. The P.E. of a correlation of -0.22 or -0.09 is so great, that the above regression equations and tables must be regarded as tentative until further research on a larger scale makes the determination more accurate. We can be quite confident that there is a small negative correlation between score and gain due to practice. Otherwise the above regression equations and tables are of value only as illustrating method.

Published norms are presumably obtained from first givings of tests. We need norms for givings of a test other than the

TABLE XII. AGE NORMS FOR THE NATIONAL INTELLIGENCE TEST ON THE FIRST GIVING AND COMPOSITE OF FIRST AND SECOND GIVINGS

Age	First Giving		First and Second Giving	
	Scale A	Scale B	Scale A	Scale B
8 (8-8.9)	46	53	104	114
9	61	68	133	144
10	76	80	162	167
11	93	96	195	198
12	109	110	225	226
13	124	123	254	252
14	133	132	272	269
15	136	135	277	275

first. The writer proposes tentatively the norms of Tables XII and XIII for the N.I.T.

In closing let me warn the reader that the accuracy of a test is not a criterion for its use. One should choose a test because of its accuracy only when other criteria are equal. The validity

TABLE XIII. GRADE NORMS FOR THE NATIONAL INTELLIGENCE TEST ON THE FIRST GIVING AND COMPOSITE OF FIRST AND SECOND GIVING

Grade	First Giving		First and Second Giving	
	Scale A	Scale B	Scale A	Scale B
III, Low	36	42	85	92
High	49	57	110	122
IV, Low	63	69	137	145
High	69	75	148	157
V, Low	84	87	177	181
High	89	91	187	189
VI, Low	102	103	211	212
High	105	104	217	214
VII, Low	117	116	241	238
High	124	121	254	248
VIII, Low	129	126	264	257
High	133	129	272	263

of a test, which is its correlation with some objective criterion, is the basis for choosing a test.

The point which it is the purpose of this paper to stress is that neither the P.E. of a score nor the gain that comes from repeating the test in another form is constant at all score levels. The P.E. of a score increases with the score. This usually implies that the accuracy for grade placement decreases in the upper grades especially since the difference in score between successive grades usually decreases in the upper grades also. The gain which may be ascribed to practice effect decreases as the score increases. The actual $P.E._{score}$ and gain in score on repetition of a test may be determined for all levels in the test by the regression equation. Those refinements in our knowledge of the tests ought to make the test much more accurate instruments in the hands of the school administrator. Of course, we should make our tests with as high reliability as possible, but the choice

of a test does not depend primarily on its reliability. The reliability of a test does, however, tell one how long a test to give or how often to repeat a test in order to obtain results coming up to a certain desired accuracy. The purpose of this paper has been to state this reliability in terms easily comprehended and used by the school administrator.

BIBLIOGRAPHY

Gates, A. I. "Study of Depth and Rate of Comprehension in Reading by Means of a Practice Experiment." *Journal of Educational Research.* 7:37-49. January, 1923.

Kelley, T. L. "The Reliability of Test Scores." *Journal of Educational Research.* 3:370-9. May, 1921.

Otis, A. S., and Knollin, H. E. "The Reliability of the Binet Scale and of Pedagogical Scales." *Journal of Educational Research.* 4:121-42. September, 1921.

Renshaw, Samuel. "The Intelligence of Teachers in Training with a Note on Practice-Effects with Intelligence Tests." *Journal of Educational Research.* 7:28-36. January, 1923.

Thorndike, E. L. "Practice Effects in Intelligence Tests." *Journal of Experimental Psychology.* 5:101-7. April, 1922. Bibliography appended.

Editorials

FUNDAMENTALISM IN RESEARCH

Our title suggests belief in the creed of our grandfathers. We confess an admiration for this creed; but nothing of the sort is now in our minds. We are using the word fundamentalism as one would have used it before it became the antithesis of modernism. We use it as it ought to be used, to designate a tendency to or belief in that which is original and basic.

We have observed in many of the practices of educational research workers a tendency to shallowness. We have taken occasion to point out more than once a lack of sustained effort, a willingness to flit from one thing to another, and an unwillingness to stay with a problem until fundamental—the word seems to haunt us—until fundamental results are secured. The inexorable curve of learning applies. In acquiring skill in typewriting, French, billiards, or making button-holes—in any prolonged act of learning—we start out bravely. Progress is rapid and interest is high. Then come the drab days of the inevitable plateau. Our lights should teach us its meaning; yet those who should know its nature are often tempted beyond their endurance to give up and to experience the delights of beginning something new. So it is too frequently in research. The problem which seems so promising becomes unattractive; difficulties assume momentous proportions; mountains become molehills; and we hasten to draw immature "conclusions." We are threatened with becoming mere dabblers in research, foolishly confident of the virtues of a fresh start. Such trifling sustains the same relation to research that mere devices sustain to true method in teaching.

There is another way in which we frequently fall short of being fundamental. We are ever anxious to study practical questions. We should, no doubt, study these questions; but much depends on our idea of what is practical. Moreover, we should be willing to devote some attention to problems which do not lie upon the surface.

For example, in the field of instructional research, it is not sufficient to consider problems of method from the point of view of school subjects. To some extent these are secondary. The primary problems of method have to do with learning. A method of teaching speed of reading or long division or appreciation of *Macbeth* is an extremely complex derived activity. Any such method is a highly wrought synthetic adaptation of simpler and more essential modes of presentation. Unless we are acquainted with these simpler operations, our highly organized procedures may be no better than the devices of which we have spoken. For no method of teaching a school subject is invariable. Indeed, even the "best" method may be inferior for many pupils and it will become inferior for all of them if it is unchanged. An unvarying method is intolerable.

If, then, teachers would be really resourceful, they must know much about the elemental methods of learning simple materials. They must know how to combine these into flexible composites, scientifically defensible. But this is educational psychology, you will say. No matter; it is essential. And if we desire to pursue these things, it will in no wise harm us to be called educational psychologists. Besides, the educational psychologists are rather slow about it. We lately looked through a number of textbooks designed for use in teaching psychology to prospective teachers. Sensation, perception, imagination, something about intelligence testing (as an apparent afterthought) attention, the emotions—these constituted most of the material. All too little appeared in these books concerning the things now known about learning.

For example, the relatively simple work of Starch, Kirby, and Miss Munn on the distribution of practice is more important

than a treatment of infra-human intelligence. Little and often, as an effective learning program, is worth more to the teacher both as a teacher and a human being than discussions of pitch discrimination and thresholds of consciousness.

Transfer of training not as a doctrine but as a condition of learning—how far under given conditions it may probably be depended upon and how far a direct and specific attack is indicated—this topic, though grossly overlaid by much twaddle, is undoubtedly "practical."

We have children in school who, though bright, cannot learn to read by the methods we are using. Have we enough fundamental expertness to accommodate our method to the learning types of these children? A working knowledge of types—visual, auditory, motor, and their compounds—will prevent our making many grievous mistakes through overworking a "best" method. We know that learning types are hard to detect—that is, to detect with the accuracy which psychology requires. We know too that the existence of pure types is problematical. But there are in every primary class a few children who with a little testing and trial can be identified, for example, as dominantly motor-minded. The usual visual methods of teaching reading will not apply; and if the teacher is to construct a method which does apply, she will have to know much about simple learning.

We have no intention to extend our examples of important fundamental data in the field of learning. Forgetting and the distribution of recall—or shall we say reviews?—necessary to counteract it, the whole and part methods of learning, the curve of work during a particular period, the learning curve over successive periods, fatigue, motivation, the meaning of the law of age of associations, and the rôle of the teacher in controlling these processes—these are some of the fundamentals.

To such as these we urge that a share of the educational research worker's attention be given. It is due to failure in this respect that much of our investigation has lacked grounding. How otherwise this may be, is evinced by the opposite tendency in Europe. Just before the outbreak of the World

War great activity was being manifested, especially on the Continent, in experimental education. In this field many periodicals were founded, most of which have not survived the holocaust. Many outstanding men studied, wrote, and taught—Meumann and Stern in Germany, Claparède in Switzerland, De Sanctis in Italy, Binet in France, and Winch in England. Some of these men are dead, others have become inactive. Of those whom we have named, Winch alone seems to be carrying the torch.

But this activity, though called educational, was not usually such as we should so regard. It had to do with what we consider psychology, of the educational sort. Moreover, this likewise appears to have been the field of thinking of progressive classroom teachers. In those days, there were research organizations not unlike the associations which we now have and which we think of as so very *fin de siécle*. The teachers of the city of Leipzig had one. It was founded in 1906 under the leadership of Rudolph Schulze. By 1913 it occupied seven rooms and had a library and much apparatus. Its members at that time were two hundred in number and in six years it had published four good-sized volumes. But its interests were dominantly psychological. Its experiments, even when carried out in the classroom, had to do with such matters as sensation, fatigue, learning, testimony, and questionnaire studies. To us there seems to be a failure to come to grips with real teaching problems. This feeling is no doubt justified.

On the other hand—and this is what we wish to point out—there is in the American *penchant* for quick results an undoubted danger. We are fairly open to the charge of shallowness in our research undertakings and of unwillingness to go back to psychology, and especially to the problems of learning, for the fundamentals of our thinking and for the bases of our procedures.

Without forsaking the direct treatment of the more complex processes let us not forget the simple, elementary, and original processes which lie beneath the surface. Let us be fundamentalists.

Reviews and Abstracts

DOUGHERTY, MARY L. *How to Teach Phonics.* Boston, Houghton Mifflin Company, 1923. 89 pp. (Riverside Educational Monographs.)

This Riverside Monograph gives in simple well-organized form, a rather thorough presentation of the teaching of phonics. Emphasis is placed upon a suggestive course of study, methods, games, and drill devices. Extensive word lists are given with the suggestion that they be modified or adapted to suit the vocabulary and needs of the class. There is no material particularly prepared for the instruction of teachers who know little or nothing about phonics.

The course outlined may be rather heavy for the average first grade of the public schools. There is one paragraph on preliminary informal training which should precede formal phonics training. The outline for third grade briefly suggests a continuation of phonics review and application with some study of prefixes and suffixes and preliminary dictionary work. No mention is made of the application of phonics to a systematic study of syllabication, such as is sometimes offered in this grade.

This little book will prove helpful to the student or inexperienced teacher, and will offer some suggestions even to the more experienced primary teachers.

MYRTLE L. KAUFMANN

Elementary Supervisor, Logansport, Indiana

HANDSCHIN, CHARLES H. *Methods of Teaching Modern Languages.* Yonkers, World Book Company, 1923. 479 pp.

The conception of this book is excellent; the point of view is broad and the treatment of the material comprehensive. *Methods of Teaching Modern Languages* would, however, have been more effective if the author had more carefully organized the material. It is longer and harder reading than it actually needs to be. In a revision of the book, the author should also correct many mistakes that occur in the illustrative examples given in the foreign languages. The book presents the best thought that is available regarding the foundations of language teaching and the aims, values, details of method, organization, and administration of the subject. It also contains a very valuable bibliography of American contributions to methodology from 1913 to 1922. From the narrow point of view of the teacher, more space might well have been devoted to a discussion of the various elements that go to make up the foreign-language course and less to principles and general organization. It must, however, be said, in fairness to the author, that there is a great deal of valuable material

packed away under the heading, "Handy List of Teaching Devices," which follows each of the chapters devoted to reading, writing, and grammar.

Professor Handschin has done modern-language teachers a great service in writing this book. It is a most important contribution to methodology and should be studied by all those who are teaching or are interested in modern languages as a subject of instruction.

E. W. BAGSTER-COLLINS

Teachers College, Columbia University

WELLS, H. G. *The Story of a Great Schoolmaster; Being a Plain Account of the Life and Ideas of Sanderson of Oundle.* New York, Macmillan Company, 1924. 176 pp.

Sanderson of Oundle—the very words have the rush and rhythm of a fresh wind blowing over moor or meadow to the sea, a refreshing and inspiring breath from the hills. That was what Sanderson was in his school. That is the atmosphere of the story of this great schoolmaster written by a master story-teller.

It is difficult to give you the flavor of this story without quoting it because each bit in it is so perfectly set, so fittingly fine that one fears to touch it lest he mar it for the reader.

The truest portrayal of the schoolmaster is in the fifth chapter of the book where he is pictured during a scripture lesson. It is masterly characterization of a great teacher who, without pomp or pride, was actuated by a great love for the boys he taught, and was inspired by a lofty ideal of truth and the teaching of truth. Sanderson was "like a sweating, panting, burly leader pushing a way for himself through a thorny thicket."

One feels that tremendous personality throughout the book as it tells of the vision of education which grew with the growth of the master. Beginning with the old competitive notion of education in which one gathers knowledge that he may gather honors to the extinction of his neighbor, he led his school to the plane of group work where each contributed his share to the task of the moment.

One feels that this steady growth, away from the selfish individualistic standpoint toward the sharing of work and product, is fundamental to Sanderson's philosophy. He would train children toward group consciousness so that the industrial competitive system, as it now stands, would give place to one where the joy of creative work contributed to the happiness of the people and the spiritual and material welfare of the workers.

It would seem that Sanderson died all too soon. But who knows? At least his message is written here for all of us who care to read, which means every school man and woman in this country.

ANGELO PATRI

Public School No. 45, The Bronx, New York

KILPATRICK, WILLIAM H. *Source Book in the Philosophy of Education.* New York, Macmillan Company, 1923. 365 pp.

A volume such as this collection of quotations on educational and related topics, hardly permits of a critical review. Its primary purpose, given in the

Preface, is to supplement, for members of the compiler's classes in Teachers College, Columbia University, "otherwise unavailable reading resources, especially by rendering many short and inaccessible references easy of access." This promise it fulfills generously. There are gathered together between the covers of the book something over 550 quotations, long and short, from sources so many and varied that only a most thorough search of the field of educational literature could have accomplished the result. Used in connection with the topics discussed in the author's classes the book should prove a handy encyclopedia of quotations. For the same reason, any one who desires, in the form of briefer excerpts, a compilation of many of the striking things which have been said on Social Control, Education and Social Progress, The Education Process, Thinking, Moral Education, and a dozen or more other topics related to the philosophy of education, will find here material exceedingly difficult to get otherwise.

On the other hand, if it is a philosophy that is wanted, it would be a mistake to hope for assistance from this book. A very different source book in educational philosophy is, therefore, conceivable. A source book which in place of compiling hundreds of quotations lifted from their setting, and representing every available angle of approach, would bring together the chief competing educational philosophies, stated in sufficient length and with sufficient force to create a genuine conflict in the reader's mind. Such a compilation would give him an incentive to work his way out to a standpoint of his own, vital, in any case, and perhaps original.

M. C. Otto

University of Iowa

Stevenson, P. R. *Smaller Classes or Larger. A Study of the Relation of Class-Size to the Efficiency of Teaching.* Bloomington, Illinois, Public School Publishing Company, 1923. 127 pp.

Just now, when the financing of schools is such a serious problem for practically every city in the country, and when administrators and school officials are earnestly seeking a solution of the problem, any light which can be thrown upon the subject is indeed welcome. The matter of class-size has a direct bearing upon school costs. To increase or decrease the average number of pupils per class in a given school system will immediately affect the costs for that system. But will the efficiency of teaching likewise be affected? The most recent contribution to the subject of the relation of class-size to the efficiency of teaching is found in the monograph *A Study of the Relation of Class-Size to the Efficiency of Teaching*, by P. R. Stevenson, research associate, Bureau of Educational Research, Ohio State University, published by the Public School Publishing Company, Bloomington, Illinois.

In the first chapter Doctor Stevenson summarizes other similar investigations which have been made by Boyer, Bachman, Harlan, Elliott, and Breed and McCarthy, but he has gone further than any of these, since he has employed intelligence and educational tests and has utilized the principle of the control method in measuring the achievements of pupils. At the same time

he has recognized that, in arriving at any definite conclusion regarding the relation of class-size to the efficiency of teaching, such factors as the ability of teachers, the supervision of instruction, and the course of study must be taken into consideration.

In chapter II the author shows that the average size of classes throughout the country in elementary schools is 38; in junior high schools, 28; and in senior high schools, 25. He finds that as the size of the school system increases the classes become larger, but asserts that the opinion of educational experts and practitioners "leans heavily toward the small class as desirable." Replies to a questionnaire sent to superintendents in cities of 25,000 or more asking the ideal size of classes, showed that the median judgment for elementary schools was 32; for junior high, 31; and for senior high, 25.

Chapters III, IV, V, and VI deal with an actual school situation found in the city of Chicago. Twenty-five small and twenty-five large classes were studied. The improvement shown over a period of two semesters is measured and the relative advances made by the small and large groups are compared. By means of the data thus collected, Doctor Stevenson demonstrates, by as accurate a means as has yet been developed, that in the elementary school small classes do relatively better than large classes. He further states that dull pupils profit more by instruction in small classes than do average or bright pupils. He finds that pupils in the second grade do as well (or better) in large classes as in small, but that pupils in the fifth and seventh grades do relatively better in small classes. Bright pupils, he states, do relatively better in large classes.

In securing data for his study of high-school class-sizes, Doctor Stevenson uses the Terman Group Test of Mental Ability in classifying 2,821 pupils into large and small groups of equal ability. He concludes from this study that the small classes in high school do only slightly better than the large classes. The study further indicates that in the high school the small classes were more favorable to the average and dull pupils than to the bright.

Although Doctor Stevenson does not make the claim that his findings relative to class-size in the elementary and senior high schools are conclusive, nevertheless, his study marks a forward movement in educational research.

J. P. O'HERN

Assistant Superintendent of Schools, Rochester

WOODY, THOMAS. *Quaker Education in the Colony and State of New Jersey.* Philadelphia, University of Pennsylvania, 1923. 408 pp.

The history of American education is in need of many such treatises as *Quaker Education in the Colony and State of New Jersey*, where the chief body of the work is composed from original sources. Because of the abuse of the use of sources in Doctor's dissertations, there has been an unfortunate reaction against the publication of works of this character. It seems good, therefore, to find a volume in which the author has had the courage to depart somewhat from conventional present standards. The illumination of the work by pertinent charts, photographs, and facsimile documents should give it a place among the source books in this field.

No attempt has been made to gloss over the weaknesses of the religious body treated in this volume. We have not here another example of the crime which has usually been perpetrated in discussing New England Puritanism, for facts which do not glorify are given an adequate place.

The chapter dealing with the philosophy of the leaders in Quaker education is very illuminating. The false notion of the unfavorable attitude of the Quakers toward higher education is corrected. The statement made by Edward Parish at the founding of Swathmore College, "In no branch of domestic expenditure is parsimony so misplaced, in nothing is it so inexcusable as in the matter of education," indicates the attitude of the leaders. They believed in "guarded religious education" for Quaker youths, but they also held that learning was properly the right of all and that it was necessary to the most successful and enriched life.

The chapter on "Education of Inferior Races" is also noteworthy. Here the author shows the Quakers' interest in the education of the Negro and Indian races and the general policy of the officials to give them an equal opportunity with the white race.

A more adequate treatment, however, of such topics as school buildings and grounds, school support, the curriculum, teacher training, tenure and salary of teachers in their relation to economic conditions and religious beliefs, Quaker influence on public schools, the modification of their own procedure by outside influences, might have been given instead of allowing so much space to meaningless details and dates. In chapter IV, for example, 269 dates are given, and these do not include the dates of manuscripts cited and of reproduced documents. This criticism seems to be true of about 200 pages of the book, that is from the third through the sixth chapter. Had this material been presented topically rather than geographically a more significant grasp of Quaker methods of solving educational problems might have been gained. Pages are given to varying rates of tuition with but slight interpretative comment. Their censorship of books would have acquired added meaning if a list had been given of those prohibited. Such a list would have thrown considerable light upon the Quaker attitude in general.

A common historical error is committed in the saying, "Education of colonial days existed primarily for religious ends." This was true of education of a certain type, but apprenticeship education and education of the poor was largely an economic matter. Moreover, one is not justified in saying that the prevailing Latin grammar schools were religious.

A more detailed impression of the transition from local or Quaker control would have been interesting. While a chapter is devoted to this, it is not a fair historical presentation for this chapter makes it seem that the Quakers yielded quite readily to the pressure for state control; whereas the real facts show that they were comparatively helpless in their struggle against the general sentiment.

In the Preface the author "solicits criticism of a constructive nature and correction where errors of fact may occur." It is hoped that criticisms given here are of a constructive sort. The treatise casts considerable light upon the struggle between the religious schools and the movement for state control, but

does not throw sufficient light upon the nature of Quaker education and its attitude toward the kind of education which was sought by those who wished to have it provided by the commonwealth.

A. O. HANSEN

Alpine, New Jersey

COY, GENEVIEVE LENORE. *The interests, abilities, and achievements of a special class for gifted children.* New York City, Teachers College, Columbia University, 1923. 194 pp. (Teachers College Contributions to Education, No. 131.)

Doctor Coy's study of a special class for gifted children is easily the most comprehensive investigation of the characteristics and development of children of superior intelligence yet reported.

The investigation is introduced with a very careful report of the history of the development of educational procedure for the gifted and a survey of the present status of the achievements thus far promoted. The thesis is a scientific study of a class of children of superior intelligence in the public schools of Columbus, Ohio. The range of intelligence quotients is 120 to 159. Extensive experiments were made in achievement and improvement in educational and intelligence tests in comparison with a control class of average intelligence. These tests began with the Binet-Simon intelligence test with intelligence quotients taken in 1918, 1919, and 1920 for organization and continued investigation of abilities. A variety of educational tests were used. The scientific handling of the data, coupled with thoughtful consideration of variants, is an assurance of the high value of the conclusions. An argument appearing early in the discussion is the increase of the accomplishment quotient of children in the special class from .963 in February 1919 to .990 in June 1920. Doctor Coy states that this increase is especially significant "since studies by Pintner, McCall, and Franzen show that in the cases of bright children who are left in ordinary classes the educational age falls farther and farther below the mental age as the children move up through the grades. In this special class we find just the opposite to be true."

An intensive investigation with material that is difficult to reduce to scientific terms was made of improvement in English composition. It was found that children of superior intelligence write best when they are given large opportunity for "productive imagination and that they are not so susceptible to social stimuli as those are who are of lower intelligence." The superior class made an improvement of 20 months during 13 months, in comparison with the control class which made an average improvement.

The chapter on classroom activities is a delightful study of work in music, original composition, reading, nature study, excursions, and history. It is full of suggestion for general teaching as well as for special-class teaching. One of the finest sections of the chapter is that devoted to the stenographic reports of recitations in special and control classes. With an "infinite capacity for taking pains," Doctor Coy has developed studies of vocabulary and teacher and pupil activity which means the length, frequency, and quality of children's and teacher's speeches, questions asked by each child, spontaneous suggestions

for solution of difficulties, information volunteered, criticism, irrelevant speeches, and other classroom responses to situations of similar nature. The reduction of this work to statistical as well as qualitative values is one of the interesting and valuable contributions to scientific achievement. Some of the conclusions are: The special class asks more questions than the control class—15.3 percent as opposed to 2.7 percent. The special class volunteered more information and a larger percent of it was characterized by evaluation and inference. The special class is more critical and more disposed to side track the argument. As Doctor Coy remarks, the teacher is a vital factor in these activities.

Chapter VIII is devoted to an interesting case study of gifted children and chapter IX to the progress of the children in the eighth grade after they left the special class. This chapter is rather disheartening but is full of suggestion for teachers and administrators.

The book is a splendid contribution to the study of a very significant movement in modern education. It is unique among Ph.D. dissertations, since it is decidedly interesting to persons other than experts in the field. It will yield its benefits not only to the scientist but to the layman as well and will be a valuable acquisition to every teacher's working library.

HENRIETTA V. RACE

Bureau of Educational and Psychological Research, Youngstown, Ohio

News Items and Communications

This department will contain news items regarding research workers and their activities. It will also serve as a clearing house for more formal communications on similar topics, preferably of not more than five hundred words. These communications will be printed over the signatures of the authors. Address all correspondence concerning this department to Doctor E. J. Ashbaugh, Ohio State University, Columbus, Ohio.

Mr. A. S. Barr, who has been working in instructional research in the Detroit public schools, will join the staff of the University of Wisconsin next September.

The January 21 issue of the *Los Angeles Educational Research Bulletin* contains a list of professional reading for teachers of art, of commerce, and of agriculture, as well as a list of educational literature for principals and teachers in secondary schools.

Mr. D. A. Worcester, director of educational measurements and standards, Kansas State Teachers College, Emporia, Kansas, has been selected to direct a mental classification of the pupils of the Kansas State Training School for the Feeble-Minded at Winfield. Mr. Worcester began his work on this assignment about the middle of February.

Coffee County (Alabama) is using the Stanford Achievement Tests plus teachers' judgments as a basis for promotion from the elementary school to the high school. Walker County has ruled that a seventh-grade child must have had 80 percent of the school term in the seventh grade before he is allowed to take the county examination for promotion to high school.

M. E. Ligon, principal of the high school, Ashland, Kentucky, has been appointed professor of education at the University of Kentucky to succeed Professor McHenry Rhoads recently elected superintendent of public instruction of Kentucky.

Professor Ligon is a graduate of the University of Chicago. He has had a rich experience in the public schools of Kentucky as teacher, superintendent, and high-school principal.

Two mimeographed bulletins dated January and February, 1924, from the State Department of Education of New York are devoted to reading. The first bulletin attempts to define the problem and lists a selected annotated bibliography on the subject. The second presents definite suggestions to teachers on how to test and how to improve the rate of reading.

Indiana University has announced its First Annual Conference on Elementary Supervision for April 17, 1924. This will immediately precede the Twelfth Annual Conference on Educational Measurements. The two conferences are placed on adjacent dates in order to accommodate the school men who will be interested in attending both sessions.

Superintendent Albert S. Cook of Maryland reports that on the statewide survey, using the Stanford Achievement Test, the results were very encouraging. In the two reading tests the graded schools were almost at standard, while the rural schools were generally about a half year below. In arithmetic, the graded and two-room schools were above standard, and the one-room schools almost up to standard. Compared with the results obtained two years ago, remarkable improvement was shown.

Mr. H. T. Manuel, director of educational research, Western State College, Gunnison, Colorado, announces the annual high-school scholarship contest for Friday, May 9. Each school may make entries in Algebra I, Latin I, American History and Civics, Physics, English, and general information. The contest will be held at the college, and free local entertainment will be provided for visiting contestants, and one faculty representative from each group.

Due to the increasing pressure of his official duties as commissioner of education, Doctor Frank P. Graves has resigned the editorship of the *Educational Review*. Mr. William McAndrew, superintendent of the Chicago schools, becomes editor with the June issue. By his noteworthy series of articles in *World's Work* as well as by his election to the Chicago superintendency, Mr. McAndrew has come prominently before the educational public within the past year. The *Educational Review*, under his leadership, should yield increasing value to its clientèle.

Doctor P. M. Symonds, professor of education and psychology in the University of Hawaii, has prepared a correspondence course of fifteen lessons in educational measurements for the classroom teacher. Teachers may enroll in this course, do the work assigned, be examined, and receive college credit at either the University of Hawaii or the normal school. Others may take the course without examination or credit.

Miss Jessie F. Cogswell, director of educational research, South Pasadena, California, has devised a combination enrollment and cumulative record card for use in that school system. Those interested in cards of this nature should write directly to Miss Cogswell. She has been able to bring important types of information together into a convenient form for filing and research purposes.

Mr. R. C. Maston, superintendent of schools at Martins Ferry, Ohio, and Mrs. Margaret Brainerd, director of research in the same school system, have issued a mimeographed bulletin to the teachers for their direction with classified groups. Specific suggestions concerning the selection of material, assign-

ment, etc., are made in terms of the group to be taught, whether average, above average, or below. A very definite attempt is being made to adjust the teaching and curriculum offering to the ability of the children, with the aim, "Every pupil working to full capacity," kept constantly in mind.

The report of the testing program for the first semester of the current year in the Duluth public schools gives the results of measurement using the Woody-McCall Mixed Fundamentals, the Monroe Silent Reading Test, the Morrison-McCall Word List, the Wilson Language Errors Test, and in grade II-A, the Haggerty Reading Examination, Sigma I. Complete distribution tables are given for the tests as well as the central tendencies, and a comparison of the results obtained in September and January together with the standard score. The January scores were high in the Woody-McCall arithmetic, silent reading comprehension, and Haggerty reading.

The Kaingsu Provincial Government, China, has undertaken the reorganization of its educational system, modelled to a certain extent after schools in the United States. Mr. S. C. Tai, graduate student in Chicago University, has been appointed special research scholar of foreign education, and is assisting his government in the collection of a library of American educational books and pamphlets. Bulletins, courses of study, etc., sent to Mr. Tai for the use of his home government will be greatly appreciated.

Among the recent bulletins of the Bureau of Educational Research, University of Illinois, are *Measuring Teaching Efficiency* by Monroe and Clark, and *Notes on the Teaching of Latin in High Schools* by Barton, Clark, Pence, and others. The former gives a resumé of the efforts that have been made toward devising score cards for the rating of teachers, and a suggested plan for the measuring of teaching efficiency based on the achievement of pupils, and the personality, intelligence, experience, and training of the teachers. A selected annotated bibliography is included.

The increase in the number of drug addicts in the United States, in spite of the stringent laws which we have, has brought before Congress the proposal that an international conference be called to consider the possibility of restricting the production of the poppy and the coca shrub to the amount necessary for strictly medical and scientific purposes. It is understood that the League of Nations is considering calling such a conference, but the place of meeting has not been decided, nor the date determined. Leaders of the movement feel that such a conference should not meet in Switzerland, since that country did not sign the Anti-Opium Convention of 1912.

One of the most complete courses of study in reading which has come to our attention is that prepared by Miss Ruth E. Hilpert for grades I to VI of the St. Cloud (Minnesota) public schools. The volume shows intimate acquaintance with the research work that has been done in the field in recent

years. It contains an unusually large amount of specific directions to teachers concerning what to teach, how to teach, and where to find supplementary aids. The appendix contains as much material as the body of the volume and includes five different series of unstandardized silent reading tests which have been devised for use in the local schools.

Doctor Frank W. Ballou, superintendent of schools, Washington, D. C., in an address before the Women's Civic League and the Public School Association of Baltimore, summarized the requirements of a modern teaching force as follows: the modern teaching force of a city school system should be professionally prepared when appointed, should grow progressively after appointment, should be adequately paid, should serve on tenure, should be retired on pension, and should be adequately supervised. His development of that thesis was so eminently sound that the editors of the *Baltimore Bulletin of Education* printed his address in full for the benefit of the teachers and patrons of the Baltimore schools.

The place of the private school in the education of the children of a democracy has been the subject of discussion at various times. The idea that "the ultimate aim of experiments in education being carried on in private schools all over the country is to develop fundamental ideas and principles to be applied for the improvement of the public schools" was recently expressed by Mr. A. C. Ferm of the Modern School, Stelton, New York. He stated that in his school he gives the child as much freedom as possible, only guiding his endeavors; and that the development of personality, social equanimity, and common sense are stressed rather than the teaching of academic subjects. If private schools can show the way to teach children rather than subject matter, they will amply justify their existence.

Miss Louise Warr, administrative assistant of the Washington Irving High School, New York City, at a recent meeting of the New York Society for the Experimental Study of Education, presented a report on a questionnaire submitted to 50 teachers whose pupils were grouped on the basis of intelligence. In the teachers' opinions there were more advantages than disadvantages accruing from the scheme especially because of the lack of satisfactory curriculum differentiation under the existing system and the general attitude which has been developed through non-classification. Fifteen of these teachers expressed decided preference for teaching the lower group, three had no preference, while thirty-two preferred the higher group.

Professor Fred C. Ayer of the University of Washington, who is also director of research of the Seattle public schools, has brought out two mimeographed bulletins on the curriculum. The first introduces the problem, and includes an extensive description of the time distribution for the subjects and grades. The second is devoted to the adaptability of the curriculum and how it is suited to the needs of the children and to the needs of the city. This topic is treated under the general heads: (1) the scope of the curriculum;

(2) the relation of the curriculum to child life; and (3) the relation of education to the city of Seattle. The bulletin closes with a bibliography.

Superintendent M. W. Longman of Muskegon, Michigan, has made a statistical study of certain phases of the school situation in ten other cities of the state having approximately the same size as Muskegon. The material, which is in the form of a mimeographed bulletin, considers about forty individual items of significance to school administrators. The school census includes the enrollment in subprimary, elementary, junior and senior high-school grades, continuation schools, attendance for the various divisions, and parochial and private school enrollment. The valuation of districts, the relation of the assessed to the true value of property, the value of manufacturing plants, tax rates, and bonded indebtedness are also treated. In all these items much variation is revealed.

A New Boston Spelling List consists of two thousand words for use in grades I to VIII, inclusive. The vocabulary was selected from the Ayres Scale, the city of Boston and the city of Chicago lists, the Jones' investigation, Kelley's *List of Words Misspelled in the Diaries of Third-Grade Children*, Nicholson's *Speller for the Use of the Teachers of California*, and Thorndike's *The Teacher's Word Book*. The words are arranged alphabetically within the grade.

Boston also has a combined spelling and vocabulary list of 4,700 words, arranged alphabetically with a spelling list of 2,000 words indicated by a star. These lists are placed in the hands of the teachers, but the directions given are inadequate for the guidance of either teachers or pupils.

Bulletin No. 784, of the University of the State of New York is a report on *The Use of Standard Tests and Scales in the Plattsburg High School*, by J. Cayce Morrison, formerly specialist in educational measurements for the State. The pamphlet is a detailed report of the tests given, the scores gained by the pupils, and the use made of the results by the teachers and supervisory force of the school. On the basis of one year's work following the testing, the superintendent states that: (1) Classes have been classified to as great an extent as the size of the school will permit. (2) Test results were used to point out to the teacher whether there was the proper relation between her class results and what might be expected of students of known I.Q. (3) The results have been used as an incentive to induce individual pupils to improve their work. And (4) the results were discussed in faculty meetings and have been vastly stimulating.

Mr. George E. Carrothers, assistant superintendent of schools, Cleveland, Ohio, in a recent article in *School and Society* suggests that an experiment be conducted to see whether the layman's attitude, that teachers work short hours and have an easy time, or the teachers' attitude, that their hours are long and their work burdensome, is true. He suggests that a number of schools be organized on an eight-hour basis with teachers on duty from eight until twelve

and from one until five, with the time of arrival and departure of pupils the same as at present. The extra time in the morning and afternoon should be used by the teachers for planning their work and grading papers, with the strict understanding that no school work should be performed by the teacher outside of these hours.

The suggestion is very interesting and we hope that a number of superintendents will find it possible to offer to Mr. Carrothers their cooperation for the year 1924-25. The facts, whether they lend support to one side or the other of the controversy, are worth establishing.

A recent article in the *National Vocational Guidance Association Bulletin* by C. J. Giddings, assistant director of the Department of Attendance and Vocational Guidance, Cleveland, Ohio, presents some interesting data regarding the education and vocations of children getting work permits. In Cleveland, in comparison with every one hundred children in the first grade, there are 53 ninth-graders and 14 twelfth-graders. Of the high-school graduates, less than 50 percent continue their education. Of the boys and girls who receive work permits, nearly half of the boys and one-third of the girls come from the junior high schools. Fourteen percent of the boys and 25 percent of the girls come from the senior high schools; 20 percent of the boys and 16 percent of the girls, from the elementary schools; and 14 percent of the boys and 19 percent of the girls, from the parochial schools. Forty-three percent of the 720 boys who received working permits in September entered manufacturing plants; 24 percent, some trade; and 13 percent, the Western Union. Of the 687 girls, 36 percent entered factories; 24 percent, department stores; and 11 percent, domestic service.

The Ohio State Teachers' Association has recently published a bulletin entitled, *Financing Public Education in Ohio*, in which a thorough investigation has been made of the sources of revenue and the amounts received. The study is not concerned with the distribution of expenditures. Among the outstanding points are: (1) Nine-tenths of all the money received for public education in the year 1921-22 was the income of local school districts. (2) Eight-tenths of the total amount was derived from the general property tax. (3) During the preceding ten-year period the general property tax increased 240 percent, but the increase per child enrolled was 171 percent. (4) The proportion of the total property tax devoted to school purposes increased from 35 percent to 43 percent. (5) Cities required two and one-half times as great a proportion of the total tax levy for municipal debt as school districts did for school debt. (6) Individual districts showed a variation in the amount of taxable property per child of school age of from $983 to $96,295. Local districts showed a variation in revenue received per child enrolled of from $12 to $301.

A great many other points are developed; and those who are engaged in a financial study of similar type may secure copies of the report while the supply lasts.

Superintendent M. E. Moore of Beaumont, Texas, has issued his course of study in ten parts as follows: Part I, physio-psychic basis of the course;

Part II, language, grammar, and rhetoric; Part III, mathematics; Part IV, nature study and geography; Part V, history and citizenship; Part VI, reading and literature; Part VII, writing; Part VIII, music; Part IX, physical education; and Part X, arts and crafts.

Part I brings to the teacher's attention the most important physical, mental, and social characteristics of school children. An attempt is made to put the teacher into a correct attitude toward the educative process by directing her attention to the child's characteristics and their significance in her teaching. The various parts devoted to individual subjects define objectives in each grade and present materials and suggestions for their use. The whole course is the outgrowth of an intensive study by the superintendent and his teachers and supervisors through a period of about five years.

Doctor W. W. Coxe, chief of the educational measurements division, New York State Department of Education, announces the following projects as activities and plans of the Bureau:

1. Reclassification of pupils in small rural schools
2. The influence of silent reading upon attainment in other school subjects
3. Standardization of regents' silent reading examination
4. Construction of a performance scale for measuring general intelligence
5. Construction of a language test
6. Bulletins on silent reading
7. Construction of a course in educational tests and measurements for use in normal schools
8. Preparation of a course of study for special classes
9. Surveys
10. Development of a teaching technique which will enable teachers of English to take account of individual differences in large classes
11. Development of a teaching technique to be used in classes made up of children of foreign parentage
12. Construction of standardized tests in agriculture
13. Organization of a research association in New York State

Among the new tests which have lately come to our desk the following are noted:

The Iowa Comprehension Test, a reading test for college students, consists of three pages of history, literature, and science, respectively, in which words, phrases, and clauses are underlined and numbered, and specific questions, based on the text, are placed in the margin. The task is to designate the correct answer to each question by the number of the underscored word, phrase, or clause.

A companion test, the Iowa Entrance Examination, High-School Content, is a sixteen-page folder containing 110 statements on English, 75 on mathematics, 100 on science, and 115 on history and social science. Each of these is completed by the multiple-choice procedure.

The Hillbrand Sight-Singing Test, devised by Professor E. K. Hillbrand of Dakota Wesleyan University, is published by the World Book Company, and is intended for use in grades IV, V, and VI. Its purpose is to measure the ability of a pupil in the mechanics of oral sight reading in vocal music, to indicate progress, and to diagnose difficulties. The test is individual. It makes no attempt to measure emotional or interpretative elements.

The Kepner Background Tests in Social Sciences by Tyler Kepner of the high school of Brookline, Massachusetts, is issued by the Harvard University Press. The test consists of seven exercises gathered from the fields of history, civics, economics, geography, and sociology. Multiple-choice, completion, best-reason, and matching are the different types of responses used in the test. The purpose is diagnostic and assumes to measure the extent of the factual background in the social sciences which a pupil or group of pupils brings to any one of the social-science subjects in the secondary school.

Remedial Work in Reading With Backward and Retarded Pupils

The following is a brief report of the work which was done with a small group of pupils in the high school of Washburn, Wisconsin, during the school year 1920-21. On the basis of the Kansas City Silent Reading Test, nineteen freshmen showed less than sixth-grade standard in comprehension. Measurement with the Haggerty Delta II Intelligence Test indicated a median I.Q. of 90 with a range from 75 to 107. Since it was not possible to organize special classes, a procedure of remedial work was devised to be followed in connection with regular textbook work in English, citizenship, and general science.

The remedial work may be classified under four heads:

1. With the books opened, the teacher asked a question on the paragraph in the text and the pupil read until he found the answer to the question asked. Time was called, when a majority of the class had found the answer. Usually the question was answered by one or two sentences in the paragraph, but sometimes the question was so worded that the answer could be given only by comprehending the entire paragraph. Sometimes two, or even three, questions were asked, and the pupils were required to find answers to all the questions in one reading.
2. Pupils were directed to read a paragraph and were asked to give the substance contained therein. This was found to be very much more difficult than the answering of questions.
3. Pupils were required to find the topic sentences in the paragraphs of the assignment.
4. Teachers in English did some vocabulary work, most of it in connection with the regular assignments.

One day per week was set aside for this type of work. Children were told their individual scores, the median of the class, and the standard for the eighth grade. Special effort was made to motivate real effort in developing better reading ability. Each teacher was furnished with the speed and comprehension scores of these pupils in order that she might thereby adapt her work to the need of each. After four months of this treatment, a second form of the same reading test was given to measure the gain. The average gain in comprehension for the group was 7.3 points on the scale, which is equivalent to about a year and a half of normal gain. The following table shows the results by quartile groups:

Quartile	Chronological Age	Mental Age	I. Q.	First Test	Second Test	Gain	Percent Gain
First	14-10	15-2	103	15	26.3	11.3	75
Second	14-9	14-4	102	16	23.4	7.4	46
Third	16-0	13-4	85	15	21.0	6.0	40
Fourth	15-3	12-5	80	14	16 2	2.2	16
Group	15-4	13-9	90	15	22.3	7.3	49

It will be seen that the gain made by each of the quartile groups is decreased as we proceed from the highest to the lowest mental-age level. In other words, the children most superior in mental ability profited most by the remedial work.

C. E. HULTEN

Sturgeon Bay, Wisconsin

To the Editor of the Journal of Educational Research:

In last November's issue of this journal Mr. R. L. Morton writes a review of my little book, *The Teaching of Fractions*, to which I should like to make a brief reply.

Mr. Morton takes me to task for giving pupils such fractions as 180/288 to be reduced to lowest terms, or expecting them to find the least common multiple of 36 and 45; also for giving rules of divisibility by 2, 3, 4, 5, etc. He is opposed to so much as mentioning the least common multiple and highest common factor, and says that the newer textbooks tend to eliminate all reference to such topics. Mr. Morton is, of course, entitled to his opinion, but I question his right to say that an author who does mention the L. C. M. and the H. C. F. must be "unacquainted with the contentions of recent writers in the psychology and sociology of arithmetic."

I have given some attention to the psychology and the social background of arithmetic for a number of years, and feel naturally somewhat disconcerted to find myself classed by Mr. Morton with those who dwell in outer darkness in regard to this interesting subject. I have, however, one consolation: I find myself in distinguished company. For no less a person than John Dewey must share my fate. If Mr. Morton will look into Book 6 of the *Alexander-Dewey Arithmetic*—for the contents of which America's foremost educator declares himself in the Preface jointly responsible—he will find on page 163 such problems as finding the L. C. M. of 33, 42, 66, 70, and 84; and, more serious still, on page 162: "Give a rule for finding the least common multiple"; while on page 189 of Book 4 children are asked to reduce 224/288, 115/253, 153/221, and others of that ilk to lowest terms.

Professor Thorndike, it is to be feared, will also be suspected of not knowing much about the psychology of arithmetic; for, while his books banish L. C. M. and H. C. F., rules for divisibility are given, pupils are required to reduce 42/105 to lowest terms, and, *horrible dictu*, there is an exercise on complex fractions!

The question involved here is, I believe, of some importance in our present-day educational thinking. The tendency to rule out of the arithmetic course every topic which is not directly useful to the pupil in solving life's "problems" has in my opinion gone far enough. If a narrow utilitarianism is to decide what shall and what shall not be taught, we shall hardly know where to stop. We shall end by stripping the course in arithmetic of every vestige of its scientific character. I have in my book clearly differentiated between arithmetical necessities and luxuries, and have frankly taken the position that it is not bad pedagogy to venture occasionally into the theoretical regions of arithmetic, to give our more capable pupils an opportunity to cultivate an "unpractical" interest, and let them appreciate the acquisition of knowledge for its

own sake. After all not a few of them can be interested in things purely intellectual, and they are delighted to learn that the science of arithmetic is something more than an instrument for buying pork chops and making aprons. To a good many of our pupils, I suppose, it will never be more than that. But we should not forever feel constrained to limit our teaching to those matters only which the lowest quartile of intelligence quotients can appreciate and understand.

EDWARD WILDEMAN

Shields School, Chicago

To the Editor of the Journal of Educational Research:

I have read Mr. Wildeman's reply to my review of his book, *The Teaching of Fractions*. I want to make two comments.

First, a policy can hardly be defended merely on the ground that others can be found who pursue it. Mr. Wildeman seems to feel that his point of view is entirely justified by the fact that he has been able to find examples in arithmetic textbooks similar to those which I criticized in his book. I undertook, upon the invitation of the editor of *The Journal of Educational Research*, to review Mr. Wildeman's book, not the publications of Dewey and Thorndike. I have not taken the trouble to verify the quotations from the Alexander-Dewey arithmetics and they are, therefore, accepted without challenge. And Alexander and Dewey may require that pupils find the least common multiple of 33, 42, 66, 70, and 84, if they choose, (the least common multiple of these numbers is 4,620) but to justify the practice in terms of social needs is quite another matter. So far as I know, this has not been done by either Alexander, Dewey, or Wildeman. If Mr. Wildeman derives solace from this situation, he is welcome to it. In the light of recent studies on the sociology and psychology of arithmetic, I cannot endorse it. Nor can I endorse the practice of requiring pupils to reduce 153/221 to lowest terms, since the one common factor of the numerator and denominator is 17.

Second, if the reader will turn to my review in the November 1923, issue, page 354, and compare the statements made there with Mr. Wildeman's references to, and quotations from, that review, he will note that I have not been fully and accurately quoted. I did not say that "an author who does mention the L. C. M. and the H. C. F. must be 'unacquainted with the contentions of recent writers in the psychology and sociology of arithmetic.'" I did say this: "The author's point of view can hardly be called 'modern.' He is either unacquainted with the contentions of recent writers in the psychology and sociology of arithmetic or he ignores them." And I proceeded to illustrate my point by quoting from his discussion of *highest common factor* and *reduction to lowest terms*.

I did take Mr. Wildeman to task for requiring pupils to reduce 180/288 to lowest terms; and I also mentioned 450/675 and 540/756. Note that he quotes the fraction having the smallest terms. I did take Mr. Wildeman to task for expecting pupils to find the least common multiple of 36 and 45; I also mentioned 72 and 108 as well as 48 and 80. I also took Mr. Wildeman to task for giving rules of divisibility by 2, 3, 4, and 5; and I mentioned, in addition to these, 6, 9, 10, 12, and 15, "detail which, in the reviewer's opinion, cannot be justified for grade classes."

R. L. MORTON

Ohio University

Educational Research Association

(HARRY A. GREENE, *Secretary and Editor*)

The papers which were presented at the closed meeting of the Educational Research Association held at Chicago Thursday, February 28, will appear in the columns of this department from time to time in order that they may be recorded in the files of the association and in order that members unable to attend may have the benefit of the material presented together with the related discussion.

A Method of Coordinating Results of Mental and Achievement Scales in Surveys of School Achievement and Progress[1]

The development of the technique for making surveys of school achievement which has been worked out in connection with the making of achievement surveys in several Minnesota towns and cities has involved the use or development of three important features: (1) The selection or devising of a series of achievement tests that measure practically all phases of elementary- or junior high-school work, with plenty of duplicate copies available, and yielding scores comparable from scale to scale in the different phases of school work measured as well as in the same phase of school work. (2) The development of a meaningful as well as a practically useful series of derived measures that are comparable from one phase of school work to another, together with the necessary material aids for readily determining such derived measures. (3) The invention of a series of record cards for preserving the scale scores and presenting the derived measures in the most usable form with the smallest expenditure of time.

The selection of educational scales as measuring instruments makes feasible a series of resurveys, the results of which are comparable with one another, and allow of intercomparisons of gains from year to year as well as from one phase of school work to another. This feature involved the construction of several new series of scales as well as the revision of others.

Such derived measures as the mental-age variations and grade variations are readily obtained from a sheet of mental age norms for the various educational scales and indicate the extent, in terms of scale scores or two-month units, to which each pupil is above or below the average achievement for his mental age as well as for his grade location. Such norms have been arrived at by determining the equivalent educational scale scores for each mental

[1]An abstract of a paper presented at the closed session of the Educational Research Association, Chicago, February 28, 1924.

age on the basis of the mental-test and achievement-scale scores made by several hundred pupils in the sixth, seventh, and eighth grades. From these data a chart has been constructed from which the pupils' mental ages, their educational ages in most phases of school work, and their achievement quotients may be read off from the mental-test and educational scale scores. As the probable errors of both the original and derived measures have been determined for nearly all the scales from actual cases, the reliability of practically all measures are available. From these probable errors it is a simple matter to determine the number of tests that need to be given to secure any desired degree of accuracy in the measures.

An individual record card has been devised for the permanent school files on which may be quickly entered for several successive surveys the names of the tests and scales used, the date and grade, the scores made in the tests, the variations from the pupil's mental age norms, the variations from the pupil's grade norms at the time the survey was made, and his achievement quotient in each phase of school work measured. It is of convenient size for filing and presents the data in such an arrangement that any kind of recorded data may be clearly seen or the gains readily computed from one survey to another.

A class card for each phase of school work, filled in by the transfer of a single entry from the permanent individual record card, shows graphically the range of mental ages of the pupils in the class, their relation to the mental-age norms for the upper grades, the class mental age, the range of achievement scores, their relation to the achievement norms for the upper grades, the class achievement placement, the quality of work of the pupils, the relative emphasis given to each phase of school work, the quality of work of the girls as compared with that of the boys, and the quality of work of the bright pupils as compared with that of the dull pupils.

In addition a small record card is available for the use of the teacher on which the scores are printed to indicate the norms for each mental-age and grade level. By placing a circle by the score made on each scale and drawing a few lines from the pupil's chronological age, grade location, and mental age, the card presents a graphic representation of the interrelation of the pupil's chronological age, grade location, and mental age, of the pupil's achievement in relation to the norms for his mental age and for his grade at the time the tests were given, of the quality of his work in all the phases measured, and his educational ages in all phases of school work measured.

Tabulation sheets have also been worked out for making on a single sheet a rapid summary of the mental ages, scale scores in all the tests, mental quotients, and achievement quotients in all the tests for all the upper grades of an entire school or school system even. From these data class scores may be readily computed and compared with the norms indicated on the sheet. Intercomparisons from subject to subject as well as from grade to grade may also be readily made for the school as a whole or even for the whole school system.

M. J. VAN WAGENEN

University of Minnesota

Discussion of Dr. Van Wagenen's Paper

Mr. Van Wagenen has worked out with commendable thoroughness a system of comparing and recording scores in various educational and mental tests on the basis of similar and easily comparable units, and he has constructed convenient charts for easy transmutation of scores in one test into equivalent scores in others, for the converting of achievement scores into mental age scores, and for the ready determination of any desirable mental or educational age or achievement quotient. Where these determinations must be made the charts will save much time of calculation, and they can easily be modified to incorporate the scores of new tests as they appear. The ideas involved in this scheme are not new, but they nevertheless represent the result of a great deal of experimental and technical work. It should be noted that the "Age and Quotient Chart" does not commit one to the adoption of the entire system of recording and interpreting scores to which it belongs; it is just as serviceable for any one else using the tests equated on it as it is for this specific system of coordinating the scores. Any system that seeks to simplify and bring together into a common unit the heterogeneity of test units now in existence is by so much to be commended. The common unit in the present system is really the mental-age unit, month or year, but grade units are also rather roughly given. It was the simple conception of mental age that did so much to popularize the Binet mental scale.

When we come, however, to apply the mental-age conception to educational or achievement tests a certain difficulty arises. We find ourselves, when we work out achievement quotients, a second place removed from the original data of the test, inasmuch as we must interpret the educational age in terms of mental age, which itself is an interpretation. The chances of error here become greatly increased over that of simply getting the intelligence quotient. For example, if we determine the limits of the true mental age with a prediction that has a probability of 1 to 2 of being correct, and if we do the same for the educational age, the final probability of having the right achievement quotient within the limits set for it is only 1 to 4. The more removed we get from the original data the greater the liability to error, and this liability increases in geometrical ratio.

The first thing that we should ask of any system, such as the one we are considering, is the degree of its reliability. Respecting the achievement quotient this question applies to Van Wagenen's system or to his chart in the same way only as it applies to this quotient in general, for the present chart shows what the quotient will be when certain scores and the age are given. The only criticism that we can make of the present system specifically, in this regard, is the neglect to point out the high degree of unreliability of the achievement quotient, or the failure to give some sort of unreliability measure. But there are also probable errors made in the attempt to predict from scores in one test what would be the scores in another test. Ranges and distributions are not identical and compromises to get equal steps must be made here and there. The problem though similar to, is not quite the same as, that of determining the reliability of any one test from the use of its two or more

forms. The error in transmuting scores near the central tendency of the distribution is less than is that of scores near the range limits, say the ten and the ninety percentiles. Mere mathematical measures are often unsatisfactory because of the presence of a number of factors which in the experiments are not properly controlled, such as the heterogeneity of the group. Correlations, for example, between two kinds of performance in a certain group are often given and compared with those of other groups that have been far more or far less rigorously selected and the result is that our educational literature is filled with correlations of this and that "ability" with some other, each with the requisite reliability quotients, when as a matter of fact we all know from experience that these reliability measures are found to be seriously wrong when we attempt to predict future correlations from them. Do we not need more attention to the determination or reliability measures by practical methods, as well as by statistical formulae? In the present case it would be well to know practically how well the charts would enable us to predict scores in one test from another or one quotient from another derived through a different pair of mental and achievement tests. Tables of such reliability data can easily be compiled as experience accumulates.

My impression is that the present system puts a little too much stress on the tabulating and recording of individual data before their real dependability is practically determined. Our original test data must not be taken with too much confidence for the erection on them of complicated statistical refinements, and this applies with double emphasis to transmuted scores. I find the directions for the use of the various record cards rather obscure, this being probably due to the fact that terms are used in referring to them—"sheet of norms," "score card," "classification sheet," and "distribution sheet"—which are not defined and are not found on the cards themselves. A single system of lettering or numbering, or a few simple definitions, would add greatly to the clarification of the instructions.

Finally, though time is short, it may be well to say that we need cautions in this as in every system, against the evil of tending to make all students alike. How far can a student diverge in an achievement test from the norm of his grade or age and still profit more from association with this grade or age group than from promotion or demotion? Is not the proper ideal to let each student advance regularly in something that he *can* do and to determine as early as possible what that is? Can it be shown that repeating a grade to get nearer the norm of that grade is ever justifiable? Is not every case of a child's repeating a grade for a year or a half year a case of a huge error somewhere in our system? We must be very careful in our testing schemes not to encourage the view that every child must approximate the norms of his grade or age. Nothing would do more harm to a school system already too mechanical than to inject this fallacious view, however unintentionally, into the hosts of untrained testers.

JOSEPH PETERSON

George Peabody College for Teachers

The Interpretation of Scores in Standard Tests[1]

Interpretation of the scores obtained from standard tests is one of the most difficult, as well as one of the most important features, of educational measurements. It is the relationships between the score obtained and other more intimately known facts that make a measured result significant and helpful to the teacher.

Perhaps the first consideration is that the method should be relatively simple in operation. Any method of interpretation that involves long and complicated calculations is certain to be used less frequently than it deserves. It should be possible by means of a satisfactory method to show directly the relationships among scores in the different school subjects. The physician's diagnosis is not made on the basis of temperature alone or on the basis of any other single test score. A third requirement in a method for interpreting test scores is that it should make possible a useful measure of growth over a period of months or years. It is the relationship of one measurement to a second or third measurement made after a lapse of time which tells whether growth has been taking place.

In an effort to develop a simple, pictorial scheme of interpretation which would be easily understood by persons who were relatively unfamiliar with graphs and educational measurements, the common household thermometer was chosen as a type. Each test is represented by a different thermometer, but the scale for raw scores on each thermometer is adjusted to the average standards for entering the different school grades. In this manner the grade and the location in the grade to which the pupil's score would normally entitle him are automatically indicated as soon as a given score is represented graphically. A completed chart is a singularly graphic profile or picture of the relationships among achievements in the different subjects.

It is not necessary for a specially trained statistician to calculate a large number of indices and quotients, for the teachers themselves can easily prepare their own charts and make their own diagnoses. Furthermore, the inaccuracies of the original data are perfectly evident to any one who studies a child's completed chart.

M. R. TRABUE

Bureau of Educational Research, University of North Carolina

[1]Abstract of a paper read before the closed session of the Educational Research Association, Chicago, February 28, 1924.